Prefazione

La complessità del comportamento di acquisto è tale che, se vogliamo migliorare la nostra comprensione del mercato, non possiamo adottare un solo punto di vista. Occorre invece porsi da più prospettive, unire i saperi, le metodologie, fondere ciò che ognuna di esse riesce a cogliere, in modo necessariamente parziale ma altrettanto necessariamente complementare.

La psicologia, le neuroscienze, la statistica e l'informatica, hanno infatti aperto nuove prospettive per il marketing, dal momento che hanno ormai dimostrato che:

- la decisione economica scaturisce dalla combinazione di processi controllati e automatici, cognitivi ed emotivi;
- il processo decisionale ha una importante componente relazionale, che discende dalla nostra capacità di rappresentare la mente altrui;
- Il contesto con cui si presenta un'opzione condiziona la decisione e mette dunque in discussione l'assioma economico della invarianza.

Il nostro interesse per l'evoluzione scientifica del marketing è frutto della collaborazione tra il Dipartimento di economia ed il Dipartimento di neuroscienze, col quale abbiamo condiviso numerosi esperimenti di risonanza magnetica funzionale (fMRI) (v. Lugli, 2010, 2012, 2014 e 2015). Nel marketing, che può essere considerato come un segmento operativo dell'economia comportamentale, comprendere i meccanismi che presiedono alla scelta è fondamentale per orientare le azioni dell'impresa in rapporto ad una clientela che esprime preferenze instabili e condizionabili.

Se vogliamo comprendere i meccanismi mentali a monte della scelta, non possiamo più limitare la nostra indagine alla tradizionale tecnica dell'intervista. I soggetti intervistati danno infatti risposte cognitive, mentre il loro comportamento è dominato dalla mente emotiva. Inoltre, l'intervista è soggetta a numerosi *bias*: la formulazione dei quesiti, la relazione con l'intervistatore, il contesto dell'intervista ed il *priming*. È dunque necessario affiancare i risultati dell'intervista con le risposte fisiche dei soggetti agli stimoli di marketing. Le risposte fisiche sono infatti automatiche e meno controllabili da parte dei soggetti; soprattutto, le risposte fisiche sono frutto dell'attivazione della mente emotiva, che è quella maggiormente coinvolta nella scelta. Oltre all'fMRI, abbiamo utilizzato l'*affective computing* nello studio delle risposte fisiche agli stimoli di marketing. Le espressioni facciali sono infatti uno dei principali canali di

comunicazione delle emozioni che proviamo in risposta ad uno stimolo; nel CAPITOLO 1 daremo conto di questa tecnica e del suo impiego. Il volto delle persone interessa il marketing non solo perché è un veicolo della risposta emotiva ad uno stimolo, ma anche perché permette di riconoscere il potenziale cliente in punto vendita. Riconoscere il cliente significa attivare una risposta automatica in relazione alla sua presenza in punto vendita ed in funzione di un profilo costruito sulla base del suo comportamento di acquisto presso l'insegna. Questa possibilità nasce dallo sviluppo di nuove tecniche statistiche che imitano le reti neurali . Il cervello umano apprende e memorizza costruendo connessioni neurali e cambiando la forza di queste connessioni in funzione delle informazioni che via via vengono processate. Oggi le macchine possono svolgere le funzioni che normalmente sono associate al funzionamento del cervello umano: percepire, risolvere problemi applicando la logica, imparare, interagire con l'ambiente e riconoscere il volto delle persone. Assistiamo dunque allo sviluppo di una sorta di intelligenza artificiale (AI) che è il frutto del miglioramento degli algoritmi, della grande disponibilità di dati unitamente all'enorme sviluppo nella potenza di calcolo e memorizzazione dei computer. L'intelligenza artificiale si esprime attraverso macchine in grado di apprendere elaborando una massa enorme di dati. Si tratta in sostanza di individuare tendenze, formulare previsioni e raccomandazioni, semplicemente attraverso la elaborazione dei dati e senza ricevere specifiche istruzioni al riguardo da parte del programma di calcolo. Questi algoritmi hanno la capacità di adattarsi ai nuovi dati che ricevono come input, migliorando così la loro efficacia in maniera continua.

L'intelligenza artificiale alla base del riconoscimento facciale funziona in questo modo: i pacchetti di codice creati dal programmatore vengono connessi dalla macchina con legami più o meno forti in relazione alle informazioni inserite per consentire l'apprendimento. Di conseguenza, il programmatore di una rete neurale, una volta completato il training con l'inserimento delle informazioni rilevanti, è in grado di spiegare solo in parte come il computer prende le sue decisioni. L'efficacia dell'intelligenza artificiale dipende quindi dalle informazioni utilizzate per il *training* dell'algoritmo. Il riconoscimento facciale è realizzato attraverso macchine capaci di un apprendimento avanzato; si parla infatti in questo caso di *deep learning*, vale a dire strati interconnessi di software che formano una rete neurale. La macchina elabora una massa enorme di dati attraverso strati successivi di software che permettono un apprendimento progressivo e la scoperta di caratteri via via più complessi nel *data set*. Questa rete neurale informatica può giungere ad autonome determinazioni elaborando il *data set*, ed utilizzando poi ciò che ha appreso nella elaborazione di nuovi dati per migliorare continuamente il risultato prodotto.

L'importanza del *data set* utilizzato come *input* dei sistemi neurali di calcolo si può meglio comprendere facendo riferimento ad alcune esperienze concrete. Utilizzando le foto dei parlamentari di diversi paesi, alcuni studiosi hanno dimostrato che il sistema di riconoscimento facciale proposto da alcune imprese (IBM, Microsoft, Face++) era più accurato nel riconoscere i volti di persone di carnagione chiara ed i maschi[1]. Rispondendo a queste critiche, IBM ha cambiato il *data set* utilizzato per il *training* del suo algoritmo ottenendo miglioramenti significativi[2]. Nel comparare l'efficacia dei vari algoritmi di riconoscimento facciale, occorre dunque tener conto dei diversi *data base* utilizzati per il *training*[3].

Il rapido sviluppo di sistemi di riconoscimento facciale è dovuto da un lato al miglioramento degli algoritmi statistici e, dall'altro, alla disponibilità di enormi *data base* di immagini facciali per il *training* degli algoritmi. Il miglioramento degli algoritmi ha permesso l'identificazione automatica dei volti in tempi brevissimi e, dunque, in linea con le esigenze degli utilizzatori. Il riconoscimento facciale è un sistema passivo e non invasivo di verifica dell'identità. Grosso modo, la tecnologia del riconoscimento facciale consiste in:

- una telecamera che rileva l'impronta biometrica del viso;
- un algoritmo che normalizza l'impronta biometrica in modo da rispecchiare il formato delle immagini del *data base* utilizzato per il *training*;
- un software che compara l'impronta normalizzata dell'individuo da riconoscere con le impronte normalizzate del database producendo di conseguenza la probabilità di sovrapposizione per ciascuna;

---

[1] Cfr. The Economist February 17th 2018, p.71.

[2] "IBM has responded quickly. It said it had retrained its system on a new data set for the past year, and that this had greatly improved its accuracy. When testing the new system on an updated version of the set of politicians Ms Buolamwini and Ms Gebru had used, the firm said it now achieved an error rate of 3.46% on dark-skinned female faces—a tenth of that the two researchers had found using the existing system. For light-skinned males the error rate also fell, to 0.25%", The Economist February 17th 2018, p.71.

[3] "While existing publicly-available face databases contain face images with a wide variety of poses, illumination angles, gestures, face occlusions, and illuminant colors, these images have not been adequately annotated, thus limiting their usefulness for evaluating the relative performance of face detection algorithms. For example, many of the images in existing databases are not annotated with the exact pose angles at which they were taken" (Tolba, El-Baz and El-Harby, 2008).

- Un’espansione automatica del *data base* di immagini con le impronte facciali delle persone non riconosciute.

Nel CAPITOLO 1 introdurremo il lettore sulle modalità di lettura delle emozioni di un soggetto attraverso le sue espressioni facciali che, dunque, non vanno confuse con le impronte facciali. Queste ultime sono infatti assimilabili alle impronte digitali ed attengono pertanto al riconoscimento della persona, senza alcun riferimento alle emozioni che sta provando. Certo, l’emozione deforma il volto di una persona e complica il suo riconoscimento, ma la tecnologia è oggi in grado di superare anche questo ostacolo. Le immagini rilevate dalle telecamere vengono infatti gestite da un server centralizzato che ha il compito di confrontarle e riconoscerle attraverso un complesso mix di algoritmi innovativi di Statistica Robusta pubblicati su riviste internazionali di statistica (v. Atkinson and Riani, 2000; Atkinson Riani and Cerioli, 2004; Riani Atkinson and Cerioli, 2009; Riani Perrotta and Torti, 2012; Atkinson, Cerioli, Morelli and Riani, 2015), Reti Neurali Convolutive, algoritmi robusti di support vector machines, sviluppati in MATLAB e PYTHON. Nel CAPITOLO 2 e nel CAPITOLO 3, illustreremo dunque la nostra tecnologia di riconoscimento facciale e localizzazione del cliente in punto vendita, che si articola nelle seguenti fasi :

- acquisizione del volto del cliente;
- trasformazione dell’immagine in una matrice numerica delle caratteristiche del volto del cliente;
- cancellazione delle immagini originali (per garantire la privacy del cliente);
- confronto delle matrici numeriche riferite ai volti dei clienti frequentatori del punto vendita con quelle acquisite;
- esito confronto negativo => creazione e archiviazione del profilo del nuovo cliente da sensibilizzare ad utilizzare la APP;
- esito confronto positivo => analisi dei precedenti comportamenti di acquisto e generazione azioni di marketing in tempo reale.

Nel CAPITOLO 2 e nel CAPITOLO 3 sottolineeremo inoltre l’importanza di un database che persegue automaticamente un processo di miglioramento delle informazioni archiviate. Ogni volta che viene catturata un’immagine con qualità migliore, la relativa matrice associata sostituisce infatti quella di qualità peggiore.

Nel CAPITOLO 4, daremo conto dei principali utilizzi delle impronte facciali; trattasi di esperienze già operative, anche se ancora in fase sperimentale.

Nel CAPITOLO 5 illustreremo i possibili utilizzi delle impronte facciali nell'*in store marketing*, sottolineando in particolare i vantaggi per gli acquirenti, l'insegna commerciale ed i fornitori. Il valore aggiunto del riconoscimento facciale e della localizzazione del cliente può essere infatti ricondotto alla possibilità di:

- offrire ai clienti promozioni rilevanti, in quanto coerenti col profilo di acquisto di ciascuno ;
- migliorare l'efficienza operativa delle insegne che, col *targeting* della clientela, possono evitare che a beneficiare dell'opportunità offerta siano anche consumatori che avrebbero comunque assunto il comportamento che si vuol promuovere;
- aumentare l'efficacia delle azioni di *in store marketing* , in quanto si tratta di realizzare azioni tarate sul profilo dei destinatari e perciò stesso capaci di suscitare un maggior tasso di risposta;
- aumentare l'efficacia della comunicazione promozionale , che non è veicolata solo prima della *shopping expedition*, ma anche all'inizio, durante e alla fine della spesa;
- sostenere la propensione all'acquisto con promozioni che si rinnovano ad ogni visita e scadono al termine della *shop expedition*, facendo così leva sull'euristica della scarsità e dell'avversione alla perdita.

Nel CAPITOLO 6 illustreremo l'esperimento che abbiamo condotto nel supermercato – laboratorio della Centrale VèGè di Milano. L'obiettivo dell'esperimento non era la misura dell'efficacia in termini di *redemption* della promozione veicolata con questa nuova tecnologia, ma solo la verifica del corretto funzionamento della tecnologia in termini di riconoscimento e posizionamento spaziale del cliente, oltre che naturalmente in termini di ricezione delle notifiche, interazione e chiusura della lista della spesa scontata col conseguente accredito degli sconti.

Nel CAPITOLO 7 illustreremo la possibile convergenza tra la tecnologia di rilevazione delle espressioni facciali e la tecnologia di rilevazione delle emozioni al fine di cogliere nuove opportunità di relazione col cliente all'interno del punto vendita.

Il CAPITOLO 8 è dedicato innanzitutto a ringraziare tutti coloro che si sono adoperati per realizzare la tecnologia. In questo capitolo, inoltre, si pone particolare attenzione a sottolineare il rispetto delle normative sulla privacy in quanto le impronte facciali dei clienti vengono conservate sotto forma di matrice numerica e non di immagine. È importante sottolineare, infine, che le informazioni che i clienti inseriscono nel momento dello scarico della APP, unitamente alle informazioni prodotte dalle telecamere e dal funzionamento della APP, sono criptate di conseguenza non possono essere intercettate.

Gianpiero Lugli
Ordinario di marketing, Dipartimento di Scienze Economiche e Aziendali, presso l'Università degli Studi di Parma

Marco Riani, PhD
Ordinario di Statistica metodologica, Dipartimento di Scienze Economiche e Aziendali, presso l'Università degli Studi di Parma
http://www.riani.it

Direttore del Centro di Ricerca Interdipartimentale di Statistica Robusta per grandi banche dati dell'Università degli Studi di Parma
http://rosa.unipr.it

CAPITOLO 1 Le espressioni facciali nel marketing

(*a cura di Gianpiero Lugli*)

È possibile misurare le emozioni per comprendere meglio il comportamento di acquisto e di consumo? Una risposta positiva a questa domanda è la premessa per migliorare la performance delle imprese. Se possiamo riconoscere e misurare la reazione fisica di un individuo ad uno stimolo, si apre infatti una nuova frontiera per il *management*. Posto che le aziende non riescono a gestire efficacemente ciò che non è misurabile, occorre uscire dallo stretto recinto dell'economia ed utilizzare la conoscenza prodotta in altre discipline; solo in questo modo è possibile misurare la reazione emotiva ad uno stimolo di marketing ed orientare di conseguenza in maniera più efficace il comportamento delle imprese.

1.1 Il modello delle emozioni basiche

Secondo il modello delle emozioni basiche (Eckman, 2003; Izard, 2007), vi sarebbe un numero finito di distinte emozioni plasmate dall'evoluzione della specie che, combinandosi, originano emozioni più complesse e sfumate. Le sei emozioni di base (felicità, paura, rabbia, tristezza, sorpresa, disgusto/disprezzo[4]) sono assimilabili ai colori di base; la loro combinazione genera numerosissime soluzioni percettive/cromatiche. Raramente sviluppiamo sentimenti puri che corrispondo a singole emozioni di base. Di norma, le emozioni si mescolano; per esempio, l'orgoglio è frutto della sovrapposizione di due emozioni: la felicità e la rabbia.

Per quanto riguarda l'incidenza della componente emotiva rispetto alla componente cognitiva nelle decisioni, i neuro scienziati ritengono che il 95%

[4] Il disgusto è l'emozione che proviamo assaggiando un alimento avariato, ovvero un alimento che non rientra nella nostra tradizione alimentare. Il disprezzo invece, è l'emozione che proviamo nei confronti di una persona che si è comportata in maniera disonesta nei nostri confronti. Le espressioni facciali con cui comunichiamo il disgusto ed il disprezzo sono molto simili e, pertanto, non faremo alcuna distinzione tra le due in questa sede.

delle nostre scelte siano originate da attività cerebrali subconscie (Zaltman, 2003); inoltre, la mente emotiva decide prima della mente cognitiva[5]. L'importanza della mente emotiva non deve tuttavia portare ad una sottovalutazione del ruolo della mente cognitiva; le migliori decisioni sono frutto infatti della collaborazione delle due componenti della nostra mente. La connessione tra mente emotiva e mente cognitiva è infatti fondamentale sia per l'appropriatezza delle nostre reazioni agli stimoli esterni che per l'apprendimento[6].

L'emozione si manifesta attraverso una risposta somatica, vale a dire un'alterazione fisiologica che interviene a seguito dell'interfacciamento fra il mondo interno e quello esterno. L'evento scatenante l'emozione genera contemporaneamente risposte neurovegetative (conduttanza della pelle, ritmo

[5] "There are three light switches on the ground-floor wall of a three-storey house. Two of the switches do nothing, but one of them controls a bulb on the second floor. When you begin, the bulb is off. You can only make one visit to the second floor. How do you work out which switch is the one that controls the light? This problem, or one equivalent to it, was presented on a computer screen to a volunteer when that volunteer pressed a button. The electrical activity of the volunteer's brain (his brainwave pattern, in common parlance) was recorded by the EEG from the button's press. Each volunteer was given 30 seconds to read the puzzle and another 60 to 90 seconds to solve it. If he had not done so in the time allotted, a hint appeared. In the case of the light-switch puzzle, the suggestion was that you turn one switch on for a while, then turn it off.

Some people worked it out; others did not. The significant point, though, was that the EEG predicted who would fall where. Those volunteers who went on to have an insight (in this case that on their one and only visit to the second floor they could use not just the light but the heat produced by a bulb as evidence of an active switch) had had different brainwave activity from those who never got it. In the right frontal cortex, a part of the brain associated with shifting mental states, there was an increase in high-frequency gamma waves (those with 47-48 cycles a second). Moreover, the difference was noticeable up to eight seconds before the volunteer realised he had found the solution. Dr Sheth thinks this may be capturing the "transformational thought" (the light-bulb moment, as it were) in action, before the brain's "owner" is consciously aware of it. The Economist, April 2009, p. 86-87.

[6] "Damasio's emotion-damaged patients were not able to learn like typical people. When Eliot made a bad investment, he recognized cognitively that it was harmful to his business and his family. However, he did not have the usual accompanying feeling, such as of harm and shame, and he did not learned to avoid the harmful behavior. Consequently, he repeatedly made bad decisions, eventually losing his job, his wife, and more. His emotional impairment manifested itself as a general lack of reasonableness and intelligence, even though he still scored above average on IQ tests and written tests of social behavior", Picard (2000).

cardiaco, regolarità della respirazione, tensione muscolare) e risposte neurofisiologiche (circuito della ricompensa e circuito della punizione). Più che attraverso uno stato emotivo, è dunque opportuno rappresentare i nostri sentimenti sotto forma di un processo che si articola in due fasi. Le risposte emotive sono infatti automatiche e in un primo momento inconsce, in quanto ci accorgiamo del sentimento che stiamo provando solo in un secondo momento quando avvertiamo le manifestazioni fisiche delle emozioni. È la mente cognitiva che suscita le manifestazioni fisiche dell'emozione che stiamo vivendo predisponendoci in questo modo all'azione. Se per esempio una circostanza fuori dal nostro controllo suscita in noi l'emozione della paura, la mente cognitiva prepara il nostro corpo alla reazione più idonea ad assicurare la sopravvivenza: la fuga o il combattimento. Le emozioni hanno dunque il compito di guidare le nostre azioni sia che si tratti di gestire una minaccia che di cogliere una opportunità. Secondo Antonio Damasio[7], gli stimoli ambientali cui siamo esposti suscitano in automatico emozioni di cui non siamo consapevoli; le emozioni si trasformano poi in sensazioni fisiche di cui siamo consapevoli, che ci predispongono alla valutazione delle minacce/opportunità ed alla conseguente azione. Sempre a Damasio dobbiamo la scoperta della interazione tra mente emotiva e mente cognitiva. Pensiero e sentimento non si escludono pertanto a vicenda; è infatti necessario un input emotivo per prendere decisioni razionali; basti ricordare a questo proposito il famoso caso di Eliot, un paziente che, avendo subito un intervento chirurgico di separazione del sistema limbico dalla corteccia prefrontale, impiegava molto tempo a prendere le decisioni più banali e si rovinò finanziariamente non riuscendo a correggere scelte che sapeva sbagliate, ma che non generavano alcuna emozione negativa. La mente emotiva si attiva dunque molto prima della mente cognitiva ed invia a quest'ultima le informazioni necessarie a stimolare le reazioni fisiche che ci predispongono all'azione.

È molto raro che una nostra decisione sia basata solo su informazioni emotive; di norma, le decisioni discendono dalla interazione tra le due componenti della mente, che sono dunque entrambe attive seppur con pesi e tempi differenti. Per esempio, nella comunicazione verbale, si può constatare facilmente lo stretto rapporto fra mente emotiva e mente cognitiva. Il contenuto rappresenta, infatti, il prodotto della mente cognitiva; l'altezza, il tono, il tempo, il ritmo, le enfasi e le pause rappresentano invece una delle forme con cui si esprime la mente emotiva. Esiste dunque una collaborazione/competizione tra mente emotiva e

[7] Damasio (1995).

mente cognitiva. Soprattutto, la bontà delle nostre decisioni è frutto dell'equilibrio delle due componenti della nostra mente[8].

Il fatto che la mente emotiva si attivi molto prima della mente cognitiva implica che quest'ultima si limita spesso a confermare e razionalizzare decisioni già assunte[9]. La maggior efficienza e rapidità della mente emotiva è dovuta alla sua enorme capacità elaborativa. Secondo Timothy Wilson, la mente emotiva elabora 11 milioni di bit al secondo mentre la mente cognitiva gestisce appena 40 bit al secondo[10]; la mente emotiva lavora infatti in parallelo ed in modalità associativa, a differenza della mente cognitiva che invece lavora in serie.

Gran parte dei canali con cui comunichiamo sono inconsci; secondo A. Mehrabian, professore emerito di psicologia a UCLA, il 7% della nostra comunicazione è verbale mentre il 38% della comunicazione è realizzata con l'altezza, il tono, il tempo, il ritmo, le enfasi e le pause della voce; ciò che più interessa in questa sede è che ben il 55% della comunicazione è realizzata con le espressioni facciali ed i movimenti del corpo[11]. Le espressioni facciali sono dunque uno dei canali con cui comunichiamo le nostre emozioni, anticipando le risposte verbali che sono frutto della mente cognitiva.

1.2 Le espressioni facciali e le emozioni

Accogliendo la teoria di Ekman ed Izard, possiamo affermare che le espressioni facciali sono la manifestazione fisica immediata, spontanea ed

[8] «…emotions play an essential role in rational decision making, perception, learning, and a variety of other cognitive functions. Emotions are not limited to art, entertainment, and social interaction; they influence the very mechanism of rational thinking. We all know from experience that too much emotion can impair decision making, but the new scientific evidence is that too little emotion can impair decision making. … Today's evidence indicate that a healthy balance of emotions is integral to intelligence, and to creative and flexible problem solving», (Picard 2000).

[9] In un esperimento di *brain imaging* è stato chiesto ai partecipanti di spingere un bottone con la mano destra o sinistra, ma prima di farlo era richiesta la comunicazione verbale dell'atto. La registrazione di attivazioni neuronali molto prima della risposta verbale ha fatto concludere ai neuro scienziati che decidiamo con la mente emotiva e solo successivamente confermiamo le nostre scelte con la mente cognitiva.

[10] Wilson (2005).

[11] Van Praet (2012).

involontaria, delle emozioni che proviamo; di conseguenza, le specificità etniche e culturali non incidono sulla tipologia delle espressioni facciali, ma solo sulla loro intensità. Esiste infatti una sorta di invariabilità culturale, che Ekman e Friesen hanno verificato con soggetti appartenenti a diverse culture rispetto ad espressioni corrispondenti alle sei emozioni base. L'origine biologica delle emozioni è peraltro dimostrata dal fatto che un individuo nato cieco, e per ciò stesso incapace di apprendere la comunicazione non verbale attraverso l'imitazione, mostra le stesse espressioni facciali dei vedenti a fronte di uno stesso stimolo.

La nostra capacità di controllo delle espressioni facciali è molto minore del controllo che esercitiamo sulle parole che pronunciamo; le espressioni facciali sono infatti una reazione spontanea della mente emotiva mentre il comportamento verbale è sempre frutto di una mediazione cognitiva. Le risposte emotive ad uno stimolo, che comunichiamo attraverso le espressioni facciali, possono essere dunque contenute o alterate solo attraverso uno sforzo cognitivo che è tuttavia per definizione scarsamente efficace in quanto i tempi di reazione sono sostanzialmente diversi. La risposta emotiva è automatica (spontanea) e rapida, mentre la risposta cognitiva è subordinata ad un processo logico e quindi molto più lenta; questo significa che, quand'anche controlliamo con successo le nostre espressioni facciali per nascondere le emozioni che proviamo, interveniamo in ritardo.

Le espressioni facciali sono solo uno dei canali attraverso cui è possibile misurare l'impatto emotivo di uno stimolo di marketing. La nuova frontiera delle misurazioni biometriche per le ricerche di marketing si articola infatti in diverse tecniche:

- L'elettroencefalogramma (EEG);
- La conduttanza galvanica;
- L'*eye tracking* e la pupillometria;
- La risonanza magnetica funzionale (fMRI);
- Il riconoscimento computerizzato delle microespressioni facciali (*affective computing*).

Tutte queste tecniche si propongono di registrare il tipo di emozioni, il livello di coinvolgimento (engagement) e di memorizzazione dell'individuo esposto ad uno stimolo di marketing. Nel caso di stimoli che hanno una durata, come un messaggio pubblicitario, è importante registrare le reazioni neuronali e somatiche dei diversi contenuti nel corso del loro svolgimento; questa conoscenza permette infatti di individuare i contenuti con maggior risonanza emotiva e gestire di conseguenza il sondaggio quando si deve produrre il messaggio in diversi format.

È ovvio che la memorizzazione di uno stimolo di marketing è molto importante per il successo dell'azione intrapresa. Che cosa stimola dunque il ricordo? Sono le emozioni che ci fanno ricordare (LeDoux, 1994). Quando uno stimolo colpisce il nostro sistema limbico ed è riconosciuto in quanto associabile ad una passata esperienza, si crea una connessione emotiva. I neuroni che "sparano" insieme si connettono in modo tale che le nostre esperienze attivano reti precostituite (Banich, 2004)

1.3 La misurazione delle emozioni

Essendo possibile misurare a priori l'impatto emozionale di diverse azioni alternative su un campione di consumatori rappresentativo del target che si vuole raggiungere, si può migliorare sia in termini di efficacia che di efficienza rispetto all'attuale approccio che consiste nel procedere per tentativi modificando le politiche aziendali nel tempo attraverso la correzione degli errori. Nuove tecniche di misurazione delle emozioni sono ora accessibili alle imprese e, quindi, si apre una sfida al miglioramento della conoscenza del comportamento di acquisto e di consumo su cui si giocherà il vantaggio competitivo nei prossimi anni.

La misura delle emozioni non interessa poi solo le imprese che vendono e distribuiscono prodotti, ma anche le imprese che vendono spazi pubblicitari. Gli editori infatti possono sostenere la domanda ed i prezzi degli spazi pubblicitari offrendo agli inserzionisti misure dell'engagement e dell'impatto emozionale dei contenuti della loro comunicazione. Nel mondo *on line*, questo nuovo approccio è già una realtà; i *click*, così come i *like* ed gli *share*, sono infatti diventati misure correnti dell'engagement e dei risultati della pubblicità on line. Lo stesso non si può dire invece per la pubblicità televisiva che detiene ancora una quota assolutamente preponderante degli investimenti in comunicazione delle imprese. Trovare nuovi metodi di misura dell'*engagement* e dell'impatto emotivo della pubblicità televisiva, prima della sua trasmissione, è fondamentale per il successo dell'impresa. Analogamente, prevedere il gradimento dei prodotti in fase di lancio è fondamentale per il successo dell'impresa. Per quanto riguarda la stima della probabilità del gradimento dei prodotti in fase di lancio, basti pensare alla possibilità di confrontare il gradimento dichiarato dopo l'assaggio o la prova col gradimento mostrato dalle espressioni facciali. La possibile incoerenza tra risposta verbale e risposta facciale è sicuramente un campanello

di allarme che dovrebbe far riflettere l'impresa sull'opportunità del lancio del nuovo prodotto.

Riuscire a leggere ed a misurare le emozioni manifestate con le espressioni facciali è poi particolarmente utile per valutare l'efficacia potenziale di un messaggio pubblicitario prima della sua messa in onda. Confrontando le emozioni evidenziate dalle espressioni facciali con le dichiarazioni verbali rilasciate in un'intervista successiva allo stimolo, unitamente all'analisi del sorriso durante la risposta verbale, è infatti possibile ridurre significativamente gli errori. Sappiamo che l'esposizione ad un messaggio pubblicitario genera nello spettatore emozioni che vengono poi corrette da una successiva elaborazione cognitiva che, a sua volta, ridimensiona le emozioni ed il legame tra queste e la marca. Per verificare questo processo di correzione che porta alla conoscenza dell'intento persuasivo del messaggio, è necessario somministrare un questionario dopo l'esposizione. I quesiti somministrati dopo la visione del filmato servono infatti per individuare:

- il legame tra la mente emotiva che genera in automatico le diverse sensazioni e la mente cognitiva che le razionalizza;
- la fonte delle diverse sensazioni legando espressioni verbali a specifiche emozioni;
- l'attivazione emotiva della sola comunicazione verbale in rapporto all'attivazione emotiva riscontrata quando le parole sono accompagnate da immagini;
- lo sviluppo di una "*persuasion knowledge*".

La comprensione del meccanismo biologico con cui si formano le decisioni, ci aiuta a rivisitare il ruolo di marketing della pubblicità. La comunicazione pubblicitaria attiva infatti da un lato il cervello consapevole creando aspettative sul consumo e, dall'altro, i neuroni specchio del circuito della ricompensa anticipando i benefici del consumo. La pubblicità si propone di stimolare nell'audience:

- Emozioni negative come la paura, il senso di colpa e il disgusto, che ci orientano ad assumere comportamenti virtuosi di astinenza dal consumo di sostanze dannose per la salute (fumo, droghe), ovvero ci spingono ad un consumo moderato (lotta all'obesità)[12];
- Emozioni negative come la paura di contrarre malattie, per stimolare il desiderio del consumo di prodotti funzionali ed aumentare in questo modo la probabilità di acquisto del brand che ha innovato il prodotto

[12] Becheur et al. (2008).

alimentare con soluzioni *free from* ovvero inserendo principi attivi farmacologici;
- Emozioni negative come la paura di non poter accedere al bene per effetto della scarsità dell'offerta, in modo da stimolare l'audience a superare le resistenze cognitive all'acquisto;
- Emozioni positive come la piacevolezza gustativa, la sensazione di sicurezza associata alla naturalità degli ingredienti e dei processi di produzione, per stimolare il desiderio del consumo ed orientare di conseguenza l'audience all'acquisto.

Le emozioni negative sono molto più forti delle emozioni positive e maggiormente efficaci nell'indurre all'azione i soggetti interessati[13]. La valenza evolutiva delle emozioni negative spiega la prontezza della risposta. Dunque, sul piano della comunicazione pubblicitaria, è più facile convincere all'azione (l'acquisto) suscitando emozioni negative. Suscitare emozioni negative serve anche per contenere gli effetti del sovraffollamento dei messaggi e della multicanalità. Infatti, è sempre più difficile catturare l'attenzione in un mercato pubblicitario sovraffollato e caratterizzato da un audience orientato allo zapping e al multitasking[14]. Le emozioni negative non sono solo più efficaci nello stimolare l'azione, ma sono anche più facili da verificare. La verifica di congruità

[13] "For both men and women, fear appear as more likely to be most effective, creating positive attitudes towards the ad and the marketer, when more elaborate processing takes place", Cotte et al. (2005).

[14] Il passaggio ad un secondo/terzo schermo durante la pubblicità è senz'altro una minaccia per molte aziende che subiscono questa evoluzione del contesto tecnologico e incontrano di conseguenza maggiori difficoltà a catturare l'attenzione del telespettatore. Tuttavia, anche questa minaccia può essere trasformata in opportunità dagli inserzionisti che disegnano il contenuto del messaggio pubblicitario in modo interattivo con contenuti offerti su altre piattaforme, cellulare, tablet, personal computer. A tal scopo è possibile utilizzare tutti i social network disponibili (Facebook, Twitter, Four Square, ecc.) oltre che i canali di condivisione di contenuti (YouTube) piuttosto che un sito appositamente creato. Un esempio ben riuscito è quello di Ferrero che con il brand Estathè ha creato, in collaborazione con il "Gambero Rosso", la guida dedicata allo street food. Si trattava di un gioco che permette di vincere la guida culinaria, oltre ad una APP che permetteva di geolocalizzare tutti i locali censiti sulla mappa. (http://www.estathe.it/home/). Posto che il multiscreen è già una realtà e in futuro aumenteranno i telespettatori con questo approccio, è fondamentale che la pubblicità venga pensata per il nuovo contesto tecnologico offrendo contenuti coordinati su diverse piattaforme tecnologiche. Solo in questo caso il second screen aumenta il ROI della pubblicità e la posizione della Nielsen secondo cui "…additional exposure across screens is having a big impact on ad effectiveness" può essere condivisa.

tra emozioni che il messaggio vuole suscitare ed emozioni che l'audience ha realmente sviluppato è più facile nel caso delle emozioni negative che, in genere, suscitano un tasso di eccitazione (*arousal*) molto maggiore.

In letteratura, si ritiene che esista uno scostamento tra le emozioni che la pubblicità vuole suscitare e le emozioni realmente sviluppate nell'audience. I consumatori sono, infatti, riceventi attivi del messaggio pubblicitario, che viene "letto" in modo diverso a seconda del genere, in rapporto all'esperienza emotiva e alla capacità cognitiva di ciascuno. Individuare le emozioni suscitate dal messaggio e misurarne l'intensità è dunque una nuova via per valutare l'efficacia della comunicazione a monte della sua messa in onda.

Per catturare l'esperienza emotiva delle persone nel loro comportamento di acquisto e consumo, non ci si può basare solo sulla tecnica dell'intervista. Le emozioni non sono categorie discrete e stabili; i sentimenti che proviamo in risposta ad uno stimolo sono infatti spesso sovrapposti e mutano continuamente nel corso dell'esperienza. Non è dunque facile esprimere le sensazioni che proviamo. Il resoconto della nostra esperienza emotiva può essere inoltre pesantemente influenzato dal contesto in cui avviene l'intervista, dal modo con cui vengono formulate le domande, dalla nostra propensione alla acquiescenza fornendo risposte socialmente desiderabili, oltre che dalla recenza e dall'effetto *priming*. Il consumatore poi non può rispondere alle nostre domande sui veri motivi che lo hanno portato ad acquistare una determinata marca, semplicemente perché non conosce le motivazioni psicologiche profonde che guidano il suo comportamento. Di norma, i consumatori intervistati sulle ragioni che li hanno indotti ad acquistare una data marca rispondono con affermazioni applicabili a tutte le marche che compongono la categoria; le caratteristiche di base ed i benefici ricercati nell'acquisto di una categoria sono infatti consapevoli. Al contrario, i benefici impliciti che determinano la scelta di una marca all'interno della categoria non sono consapevoli. È ovvio che le imprese competono per differenziare il loro prodotto sul piano dei benefici inconsapevoli, perché è su questo piano che si decide l'acquisto della marca. Per migliorare la conoscenza del consumatore, sono state dunque sviluppate metodologie che registrano l'impatto fisico di determinati stimoli. L'idea di fondo è che il corpo non può mentire e, di conseguenza, le misure biometriche possono arricchire gli strumenti tradizionalmente utilizzati per analizzare il comportamento del consumatore. Oggi è possibile raccogliere dati sulla reazione del soggetto ad uno stimolo in modo del tutto passivo, cioè senza bisogno che l'individuo si impegni in relazione allo scopo. Naturalmente, quando emerge uno scostamento tra risposte verbali e risposte biometriche, conviene utilizzare quest'ultime nel processo decisionale; le reazioni fisiche sono infatti inconsapevoli, difficilmente

controllabili, e per ciò stesso più affidabili delle risposte verbali che scontano la mediazione cognitiva.

È tuttavia appena il caso di ribadire che queste nuove tecniche di indagine non sostituiscono le vecchie, ma le completano. Infatti, una piena comprensione del comportamento umano richiede anche la conoscenza della costruzione individuale e sociale della realtà, che si può ottenere solo con le tecniche dell'intervista. La mediazione culturale, sociale e relazionale, che determina le risposte dell'intervistato spiega infatti i comportamenti in cui prevale la componente cognitiva della mente. D'altra parte, la maggior affidabilità delle misure neuropsicologiche, che discende dall'insensibilità a fattori esterni e dalla minor possibilità di manipolazione intenzionale, è comunque temperata da una ridotta validità ecologica. Le risposte ottenute in laboratorio non possono infatti tener conto di tutte le variabili che determinano i comportamenti individuali e collettivi in ambienti reali.

Delle varie tecniche di misura delle emozioni provate in risposta ad uno stimolo, in questa sede interessa solo l'*affective computing,* in quando il nostro interesse è centrato sull'utilizzo del volto nella ricerca e nella pratica di marketing. L'approccio utilizzato dagli studiosi di *affective computing* è decisamente interdisciplinare in quanto occorre coniugare conoscenze psicologiche, statistiche e neurologiche con la scienza del calcolo elettronico.[15] Vi sono due diversi indirizzi di ricerca. Da un lato vi sono studiosi che si occupano della costruzione di macchine in grado di interagire con gli utenti in funzione delle emozioni manifestate e, dall'altro, vi sono ricercatori che utilizzano il computer per scoprire le emozioni sviluppate dalle persone esposte ad uno stimolo. Il primo approccio può interessare il marketing nella misura in cui il contesto viene adattato alle caratteristiche della persona e alle emozioni che sta provando in un dato momento. Si pensi per esempio ai manifesti elettronici che adattano il contenuto della pubblicità al sesso, all'età e alle emozioni dei passanti semplicemente leggendo le loro espressioni facciali[16]. Si pensi ancora alla possibilità di inviare messaggi personalizzati sul cellulare dei passanti via WiFi. Infine, si pensi alla possibilità di misurare l'efficacia della comunicazione

[15] "This book proposes that we give computers the ability to recognize, express, and in some cases, have emotions……affective computing also includes many other things, such as giving a computer the ability to recognize and express emotions, developing its ability to respond intelligently to human emotion, and enabling it to regulate and utilize its emotion", Picard (2000).

[16] The Economist January 26th 2013, p. 58.

esterna contando le interazioni col contenuto del manifesto[17]. Tesco sta sperimentando nelle sue stazioni di servizio un sistema di lettura delle espressioni facciali delle persone in coda in modo da adattare il contenuto del messaggio pubblicitario al genere, all'età ed allo stato emotivo[18].

Un'applicazione di grande potenziale è il riconoscimento facciale come strumento di pagamento; la prima applicazione è stata testata con successo in giugno 2013 nell'area di Helsinki. L'interesse di questo sistema è riconducibile alla velocizzazione della transazione ed alla possibilità di pagare senza l'ausilio di strumenti aggiunti (denaro, carta di credito-debito, cellulare) e, dunque, in completa sicurezza[19].

Pur non sottovalutando l'importanza della costruzione di macchine che adattano il contenuto alla persona che sta davanti, in questa sede interessa principalmente il secondo approccio, vale a dire la possibilità di utilizzare il

---

[17] "In Britain about 20% of outdoor ad revenue comes from digital screens. America is behind: only 1% of roadside signs are digital. (This may be because American roads are longer and less crowded, so pricey signs are harder to justify.) Measuring outdoor advertising's effect on sales, long difficult, is becoming easier. In February Postar, a market-research firm, will launch Britain's first system that tallies audiences for all forms of out-of-home advertising", The Economist January 26th 2013, p. 58.

[18] http://www.thegrocer.co.uk/companies/supermarkets/tesco/tesco-turns-to-face-scanning-ad-screens/351181.article?qsearch=1&qkeyword=screens%2f351181.article&redirCanon=1

[19] "We decided to take convenience to a whole new level. Imagine going in to a kiosk, picking up a newspaper at the cashier, clicking "Ok" on a Uniqul tablet and walking away. Imagine a day at the beach when you don't have to think about where to store your wallet during your swims. Imagine a new form of payments.

We are using military grade algorithms to make sure that the security of our system is impeccable. We have written and improved our program so that it can perform everything in the blink of an eye. We have developed an user friendly interface and made the system easy to use. All of this comes down to a user experience which consists of a user merely clicking "Ok" on our tablet. In the background our algorithms are processing your biometrical data to find your account in our database as you are approaching the cashier. The whole transaction will be done in less than 5 seconds – the time it usually takes you to pull out your wallet. We believe that we have developed the most secure and convenient payment system available and we have managed this while also bringing down the transaction time from the average of ca. 30 seconds to less than 5 seconds. You can use all the major credit cards when registering a payment method on your Uniqul account", English Finnish News, July 15th, 2013.

computer ed i relativi algoritmi statistici per scoprire le emozioni delle persone esposte ad uno stimolo.

Tutti noi cerchiamo, in diversa misura oltre che per diversi scopi e con diversa efficacia, di nascondere le emozioni che proviamo[20]. Tuttavia, posto che gli sforzi di nascondimento delle emozioni non possono essere per definizione efficaci al 100%, si pone la questione di quale sia il metodo di misura più efficace per la rilevazione e misura degli stati emotivi. È ovvio che i comportamenti meno controllabili sono quelli maggiormente in grado di esprimere le reali sensazioni dei soggetti esposti ad uno stimolo: nella letteratura che si è occupata della comunicazione non verbale, si sostiene che i comportamenti meno controllabili (L*eakage hierarchy*) siano nell'ordine le attivazioni cerebrali rilevabili con la fMRI, la dilatazione delle pupille, il tono della voce e le espressioni facciali (Ekman and Friesen 1969).

Il nascondimento delle emozioni realizzato governando le espressioni facciali può essere tuttavia efficace solo nelle relazioni interpersonali, ma difficilmente può ingannare gli algoritmi statistici dal momento che il software può leggere anche espressioni facciali che durano frazioni di secondo. L'intervento cognitivo di mascheramento dell'emozione governando l'espressione facciale è per definizione molto più lento della manifestazione fisica; aumentando dunque esponenzialmente la velocità di rilevazione delle espressioni facciali si possono scoprire sentimenti che successivamente l'individuo riuscirà a nascondere. L'*affective computing* è dunque una sorta di *WikiLeaks* delle emozioni. In particolare poi, è possibile segmentare lo stimolo incrociando la variabile temporale con le emozioni suscitate; in questo modo, si possono individuare i contenuti che generano una maggior risposta emotiva. Ciascuna emozione è inoltre codificabile per intensità utilizzando in proposito sia la scala alfabetica di Eckman sia indicatori generati dall'*affective computing*. Nel caso di un messaggio pubblicitario, si può calcolare l'*engagement* (risonanza emotiva totale), la *valence* (prevalenza del segno positivo/negativo nella risposta emotiva ad uno stimolo), l'*arousal* (intensità della risposta per ciascuna delle 6 emozioni in una scala da 0 a 1), il contagio emotivo (livello di sovrapposizione delle emozioni espresse dal volto del testimonial e dal volto dell'*audience*).

L'*affective computing* può essere dunque molto utile per la selezione del messaggio pubblicitario in grado di catturare meglio l'attenzione, sviluppare le emozioni target e memorizzare il messaggio. Confrontando le emozioni evidenziate dalle espressioni facciali con le dichiarazioni verbali rilasciate in un'intervista successiva allo stimolo, unitamente all'analisi del sorriso durante la

[20]Le persone anziane riescono a nascondere meglio le emozioni negative, mentre i giovani sono più efficaci nel nascondere le emozioni positive (Blanchard-Fields and Coats, 2008).

risposta verbale, è infatti possibile ridurre significativamente gli errori di comunicazione (Lugli, 2014).

Le distanze biometriche del volto possono essere utilizzate per compilare una matrice che ha la stessa natura delle impronte digitali; parliamo infatti in questo caso di impronte facciali. Espressioni ed impronte facciali sono dunque cose diverse (v. Figura 1 e Figura 2).

Figura 1 Esempio di impronta facciale

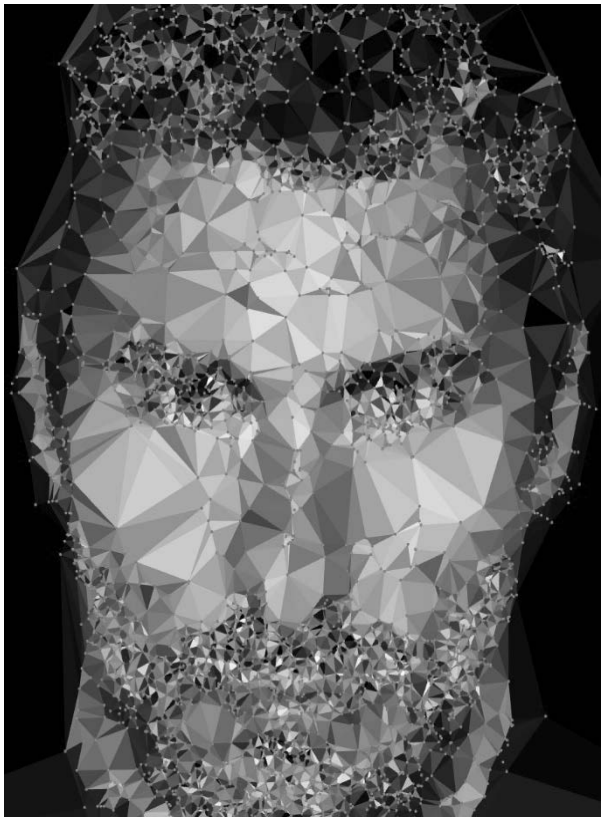


Figura 2 Esempio di espressioni facciali che sottendono le diverse emozioni di base

Come abbiamo visto, le espressioni facciali esprimono le emozioni che proviamo quando siamo esposti ad uno stimolo; si tratta dunque di una deformazione del volto rispetto alla condizione di *benchmark* rappresentata dalla situazione in cui non proviamo emozioni perché non siamo esposti ad alcun stimolo. Le impronte facciali, che vengono rilevate utilizzando telecamere ed una serie di algoritmi di intelligenza artificiale, possono riguardare sia un volto privo di emozioni che un volto deformato dalle emozioni. In quest'ultimo caso, occorrono parecchie immagini per riconoscere con certezza la persona che si trova in quel luogo/momento, agganciando di conseguenza le informazioni utili al *targeting*. Le impronte facciali possono essere utilizzate per riconoscere una persona e garantire l'accesso sicuro ad un *device* o ad un luogo, oltre che per pagare un conto e per localizzare la persona in modo da veicolare notifiche personalizzare e georeferenziate. Nei successivi due capitoli di questo testo approfondiamo il tema delle impronte facciali per concludere poi, alla fine di questo lavoro, sulla possibile convergenza di espressioni e impronte facciali come strumento innovativo di supporto all'*in store marketing*.

CAPITOLO 2 La tecnologia di rilevazione delle impronte facciali

(*a cura di Gianluca Morelli*)

L'identificazione delle impronte facciali vede coinvolta nel suo articolato processo di funzionamento una serie di metodologie note come intelligenza artificiale.

L'intelligenza artificiale è una disciplina che negli ultimi anni ha portato un importante contributo al progresso delle scienze informatiche. Tale progresso si è rapidamente trasferito nella vita di gran parte dell'umanità sotto forma di semplificazione e risoluzione di problemi quotidiani.

Alcune delle applicazioni più comuni di intelligenza artificiale sono i motori di traduzione automatica di alcuni siti web, i suggerimenti durante la scrittura di un messaggio sul proprio smartphone, gli assistenti virtuali (Siri, Google Now e Cortana), i sistemi di supporto alla guida (come il controllo elettronico della trazione che consente di avere sempre la motricità ottimale delle ruote di un veicolo), i progetti in via di perfezionamento sulla guida automatizzata e i servizi clienti di molti operatori telefonici. Altri campi di applicazione dell'intelligenza artificiale sono quelli relativi alla sicurezza del cittadino, ad esempio l'antifrode (programma THESEUS sviluppato dal Joint Research Centre della Commissione Europea[21]) e l'anticrimine.

2.1 Intelligenza artificiale

L'Intelligenza Artificiale (AI) è un concetto astratto che ha come oggetto lo studio e lo sviluppo di metodi di calcolo, basati su fondamenti teorici di natura matematica, statistica e probabilistica, che codificati in software consentono all'elaboratore elettronico di compiere operazioni di dominio dell'intelligenza umana.

In altre parole, con l'espressione intelligenza artificiale si intendono tutte quelle tecniche che rendono abile un computer a imitare l'intelligenza umana usando la logica, i processi decisionali ad albero, i legami di causa effetto e l'autoapprendimento (Figura 3).

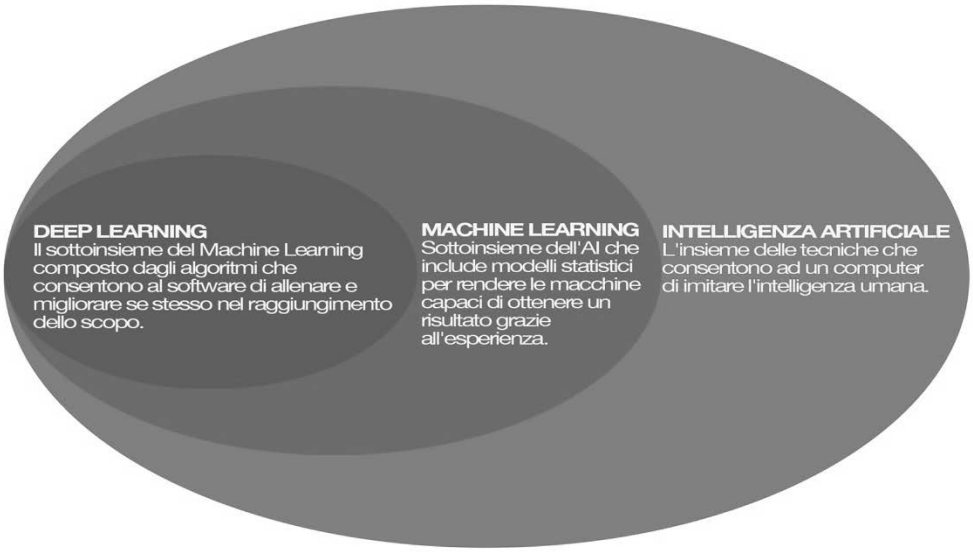
[21] https://theseus.jrc.ec.europa.eu/

L'apprendimento automatico, detto Machine Learning (ML), è lo strumento fondamentale che consente ai calcolatori elettronici di raggiungere l'intelligenza artificiale . Il processo di apprendimento automatico è un sottoinsieme dell'intelligenza artificiale che permette alle macchine di ricevere una serie di dati e di modificare in autonomia le elaborazioni sulla base delle nuove informazioni ricevute modificando gli algoritmi utilizzati. Più tecnicamente, con il termine machine learning si definisce la capacità di una macchina di apprendere senza essere programmata esplicitamente. L'apprendimento automatico è dunque un processo per educare un algoritmo in modo che possa apprendere dai nuovi dati e dai risultati passati, permettendo alla macchina di acquisire esperienza nel processo decisionale. Il machine learning usa metodi statistici (generalmente basati su reti neurali) e modelli ottimizzati per trovare informazioni nascoste o latenti nei dati.

Il raggiungimento dell'apprendimento automatico è ottenuto attraverso l'utilizzo di un approccio chiamato apprendimento approfondito o *Deep Learning* (DL)[22]. L'apprendimento approfondito utilizza modelli di reti neurali con un altissimo numero di strati (*layer*) e utilizza una enorme quantità di dati al fine di allenarsi nell'apprendere modelli complessi. Il computer, non si limita più ad utilizzare lo schema logico originario per risolvere un dato problema, ma fa tesoro del risultato e utilizza nel problema successivo l'esperienza acquisita. L'efficacia del machine learning è strettamente legata alla disponibilità di dati per educare gli algoritmi e proprio grazie all'esplosione dei Big Data degli ultimi anni le tecniche di apprendimento automatico hanno avuto una vertiginosa accelerazione nel loro sviluppo. Questa relazione non è a senso unico, infatti, come l'autoapprendimento ha bisogno di grandi quantità di dati, anche i Big Data spesso hanno bisogno dell'intelligenza artificiale per essere analizzati.

[22] Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction. These methods have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object detection and many other domains such as drug discovery and genomics. Deep learning discovers intricate structure in large data sets by using the backpropagation algorithm to indicate how a machine should change its internal parameters that are used to compute the representation in each layer from the representation in the previous layer. Deep convolutional nets have brought about breakthroughs in processing images, video, speech and audio, whereas recurrent nets have shone light on sequential data such as text and speech (LeCun, Bengio and Hinton, 2015).

Figura 3 Ecosistema tecnologico dell'intelligenza artificiale



Appare immediatamente evidente come l'intelligenza artificiale sia il risultato dell'interazione di diverse scienze quali la matematica, la statistica, la probabilità, l'informatica e l'ingegneria.

Il riconoscimento delle impronte facciali si colloca esattamente nel contesto dell'intelligenza artificiale presentando delle peculiarità di complessità di assoluto rilievo rispetto alla maggior parte delle altre applicazioni di machine learning.

La difficoltà legata al riconoscimento delle caratteristiche di un volto è determinata da una molteplicità di fattori che rendono la sfida statistica e tecnologica impegnativa e che devono essere gestiti al meglio al fine di ottenere risultati ottimali in uno degli ambiti di ricerca di maggiore fascino degli ultimi anni. I principali elementi critici per il corretto riconoscimento di un volto, a parità di efficacia dell'algoritmo sottostante, sono le somiglianze di alcuni soggetti che in presenza di una numerica di persone molto ampia aumentano di frequenza. Altre criticità derivano dai problemi legati alla varietà di espressioni che ciascun soggetto può assumere, dal look dell'individuo (ad esempio la presenza di occhiali da vista o da sole), la barba o il make-up degli occhi e delle labbra più o meno marcato. Molto banalmente, anche la semplice illuminazione dell'ambiente (Xiaoyang and Triggs, 2010) può condurre a un mancato riconoscimento, in quanto le ombre del viso possono portare ad una errata mappatura dei tratti somatici.

I fattori appena elencati sono solo una piccola parte delle criticità potenziali che, anche se non sono rilevanti per le capacità umane, rappresentano ostacoli

decisamente significativi per una macchina, considerato che devono essere gestiti in totale autonomia.

2.2 Il processo del riconoscimento facciale

Il riconoscimento facciale automatizzato, come gli altri ambiti di impiego del machine learning, necessita della creazione di classificatori, ossia algoritmi in grado di imparare aspetti della realtà che ci circonda e di compiere decisioni adeguate una volta che ci troviamo in presenza di nuovi stimoli. Il compito dei metodi di classificazione è quello di raggruppare, confrontare ed identificare tutto ciò che all'interno di un insieme di osservazioni appare uguale, simile o diverso.

Lo sviluppo di un efficiente metodo di classificazione nel caso del riconoscimento facciale automatizzato consente di individuare, dati due insiemi di fotografie o filmati contenenti volti, quanti di questi compaiono in entrambi gli insiemi e di definire con quale probabilità il confronto è esatto.

Il processo di riconoscimento facciale, indipendentemente dalle tecniche utilizzate segue sempre il medesimo schema generale (v. Figura 4). Il processo è articolato 5 fasi principali che, a seconda del metodo utilizzato, possono essere ulteriormente aggregate o disaggregate. Tali fasi possono essere sintetizzate come segue:

1. cattura ed invio dell'immagine ad un computer da parte di una fotocamera o di una videocamera connessa ad una rete dati;
2. ricezione dell'immagine da parte di un calcolatore elettronico dove un algoritmo verifica se il fotogramma contiene dei volti. In caso positivo un algoritmo avvia una prima elaborazione dell'immagine isolando i volti e li invia all'algoritmo successivo;
3. estrazione delle caratteristiche dei volti e loro trasformazione in matrici numeriche pronte per l'identificazione;
4. comparazione delle caratteristiche estratte dall'immagine con le caratteristiche di tutti i volti contenuti in un database;
5. identificazione, se l'algoritmo di confronto trova all'interno del database delle matrici numeriche identiche o fortemente somiglianti con quella in oggetto e restituzione della probabilità di corretta identificazione.

I modelli di riconoscimento facciale attualmente più efficaci hanno una probabilità di corretta identificazione che supera il 99%. Di seguito analizziamo nel dettaglio le 5 fasi schematizzate sopra.

Figura 4 Schema del processo di riconoscimento facciale

La prima fase è caratterizzata dall'acquisizione e dall'invio dell'immagine. Per raggiungere questo obiettivo possono essere utilizzate fotocamere, webcam e telecamere in grado di inviate le immagini catturate attraverso un protocollo di comunicazione dati di tipo WiFi.

La seconda e terza parte del processo consiste nel far elaborare le immagini da una struttura basata su una rete neurale a più strati. L'elaborazione delle immagini inserite nella rete neurale è organizzata in blocchi, i primi strati della rete hanno il compito di verificare se nell'immagine è presente un volto. In caso positivo il volto viene ritagliato e isolato dal resto del fotogramma e inviato a un blocco di strati più profondi che si occupano di rilevarne le caratteristiche. In questa fase l'algoritmo di estrazione delle caratteristiche facciali rileva i tratti distintivi del volto, quali forma, occhi, naso bocca e orecchie, ricavando la forma di ogni tratto somatico e la distanza dagli altri. Il processo avviene identificando le zone di contrasto ricavate dall'immagine e creando una mappatura in tre dimensioni del volto.

La quarta fase, la parte più variabile del processo, può o proseguire all'interno dell'ultimo blocco della rete neurale che va a confrontare le caratteristiche estratte con una serie di immagini già presenti in un database o chiamare in causa un altro modello statistico per effettuare l'identificazione. Questa seconda strada, prevede il taglio della rete nello strato relativo all'estrazione delle caratteristiche facciali per poi inserirle in algoritmi statistici di classificazione che effettuano il confronto tra ciò che è stato rilevato e i dati archiviati nel database. È da sottolineare che tipicamente ciascuno di questi processi di identificazione ha bisogno di una serie di immagini già presenti in un database al fine di poter eseguire il confronto e la seguente identificazione.

Per loro natura le reti neurali sono tanto più precise nel processo di identificazione tanto più sono allenate, dunque tante più foto dell'individuo da riconoscere sono presenti nel database tanto più è alta la probabilità di riconoscere correttamente il soggetto.

La quinta fase del processo riguarda la gestione del risultato. L'esito del processo può essere declinato in quattro casi. Se la nuova immagine viene associata correttamente a quelle relative all'individuo già censito nel database l'identificazione ha avuto successo, se l'immagine non viene associata a nessuna di quelle già presenti ma il soggetto è stato censito si ha una mancata

classificazione. Se l'immagine è associata ad una persona differente si ha un falso positivo o un'errata classificazione ed infine se l'immagine non viene associata a nessuno dei soggetti censiti, perché il volto è nuovo, si crea un nuovo profilo che servirà per identificare il nuovo individuo nelle sue successive apparizioni.

2.3 Modelli statistici per il riconoscimento

Le moderne tecniche di riconoscimento facciale, come descritto in precedenza, si basano o sull'utilizzo esclusivo delle reti neurali o su un mix di reti neurali e altri modelli statistici. Al fine di comprendere meglio come tali componenti del processo operano entreremo più nel dettaglio del funzionamento delle reti neurali, dei modelli di classificazione e dei concetti di analisi discriminante, tralasciando comunque formalismi e tecnicismi.

2.3.1 Reti neurali

Le reti neurali artificiali sono strumenti del dominio dell'apprendimento automatico e si basano su modelli matematici composti da neuroni artificiali che funzionano traendo ispirazione dalle reti neurali biologiche.

Le prime ipotesi di autoapprendimento risalgono al 1949 e furono introdotte da D. O. Hebb sulla scia dell'intuizione del neurone artificiale introdotto da W. S. McCulloch e Walter Pitts nel 1943. Nel 1958 F. Rosenblatt introduce il primo schema di rete neurale, detto percettrone, antesignano delle attuali reti neurali, per il riconoscimento e la classificazione di forme. Il modello probabilistico di Rosenblatt è mirato all'analisi, in forma matematica, di funzioni quali l'immagazzinamento delle informazioni, e della loro influenza sul riconoscimento dei pattern. Il lavoro di Rosenblatt, attraverso l'introduzione dei pesi sinaptici, costituisce il passo decisivo verso l'autoapprendimento. Nel 1974 Paul Werbos stabilì il contesto matematico per addestrare le reti multistrato oggi utilizzate, ovvero il metodo con il quale una rete viene preparata ad operare in autonomia.

Il principio di addestramento delle reti neurali multistrato è il cosidetto algoritmo di retropropagazione dell'errore introdotto nel 1986 da David E. Rumelhart, G. Hinton e R. J. Williams (Rumelhart, Hinton and Williams, 1986), che consiste in un meccanismo di apprendimento attraverso una procedura iterativa di aggiornamento della matrice dei pesi sinaptici che riduce progressivamente la distanza fra risultato noto e risultato della rete neurale. Le interazioni terminano per convergenza del processo, ovvero quando la

discrepanza tra risultato noto e risultato della rete scende al di sotto di un limite prefissato. L'addestramento di une rete neurale basata sulla retropropagazione dell'errore è caratterizzato da due diversi stadi: *forward-pass* e *backward-pass*. Nella fase *forward-pass* i vettori in input sono applicati ai nodi in ingresso con una propagazione in avanti dei segnali attraverso ciascun livello della rete. Durante questa fase i valori dei pesi sinaptici sono tutti fissati. Nella fase *backward-pass* il comportamento della rete viene confrontato con l'uscita desiderata ottenendo il segnale d'errore. L'errore viene propagato all'indietro lungo la rete e modifica i pesi sinaptici in modo da minimizzare la differenza tra il risultato attuale e il risultato noto desiderato.

Il processo di apprendimento si basa su tre paradigmi: apprendimento supervisionato, apprendimento non supervisionato e apprendimento per rinforzo.

Nel caso in cui si disponga di un insieme di dati per l'addestramento della rete che preveda esempi tipici di ingresso e relative uscite corrispondenti, si parla di *apprendimento supervisionato*. In questo caso la rete inferisce sulla relazione che lega i dati modificando i pesi della rete stessa al fine di minimizzare l'errore. Se l'addestramento ha successo, la rete ha acquisito la capacità di riconoscere la relazione incognita che lega le variabili di ingresso a quelle di uscita ed è quindi in grado di fare previsioni attendibili, anche quando non è nota a priori l'uscita.

Una rete può essere addestrata anche attraverso variabili che includono i soli dati di entrata, in questo caso si parla di *apprendimento non supervisionato.*

Gli algoritmi tentano di raggruppare i dati d'ingresso e di individuare pertanto, attraverso metodi topologici o probabilistici o statistici, degli opportuni cluster o pattern in grado di rappresentare correttamente la struttura dei dati stessi.

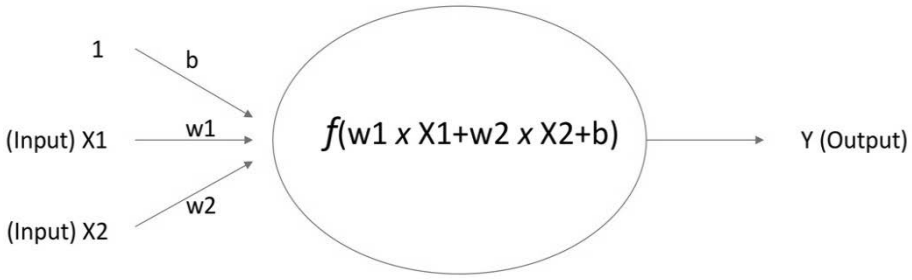
Un terzo tipo di apprendimento è il cosiddetto *apprendimento per rinforzo* il quale si riferisce ad algoritmi orientati agli obiettivi. In questo metodo di apprendimento un opportuno algoritmo si prefigge lo scopo di individuare un certo modus operandi. Tale metodo prevede un agente, dotato di capacità di percezione, che esplora un ambiente intraprendendo una serie di azioni. Ogni azione ha un impatto sull'ambiente e l'ambiente produce una risposta che guida l'algoritmo stesso nel processo d'apprendimento. L'ambiente stesso fornisce in risposta un incentivo o un disincentivo, secondo i casi. In sintesi, gli algoritmi di apprendimento per rinforzo tentano di stabilire azioni finalizzate a massimizzare gli incentivi cumulati ricevuti dall'agente nel corso della sua esplorazione del problema.

A seguito della descrizione di carattere generale appena vista si possono introdurre, sempre rimanendo su un piano informale, gli elementi fondamentali del funzionamento di una rete neurale.

Come si è visto, l'unità base di calcolo più elementare di una rete neurale è il neurone artificiale chiamato anche nodo o unità. Il neurone riceve input da alcuni altri nodi o da una fonte esterna e calcola un'uscita. Ogni input ha un peso

associato (*w*), che viene assegnato in base all'importanza relativa ad altri input. Il nodo applica una funzione *f* (definita di seguito) alla somma ponderata dei suoi ingressi come mostrato nella Figura 5.

Figura 5 Struttura del neurone

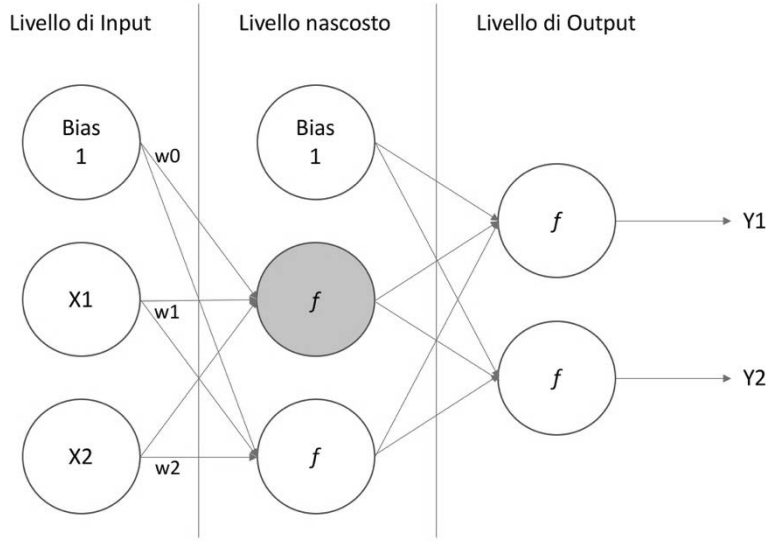


La rete di cui sopra acquisisce gli ingressi numerici *X1* e *X2* con pesi *w*1 e *w*2 associati a tali ingressi. Inoltre, vi è un altro input con peso *b* (chiamato Bias) che ha la funzione di fornire a ciascun nodo un valore costante addestrabile (oltre ai normali input ricevuti dal nodo). L'uscita dal neurone Y viene calcolata attraverso la funzione $f(w1 \times X1 + w2 \times X2 + b)$. La funzione *f* è non lineare e viene chiamata funzione di attivazione. Lo scopo della funzione di attivazione è quello di introdurre la non-linearità nell'output di un neurone. Questo è importante perché molti dati del mondo reale non sono lineari e il neurone deve quindi apprendere questo tipo di rappresentazioni. Ogni funzione di attivazione (o non-linearità) prende un singolo numero ed esegue su di essa una determinata operazione matematica. Un esempio di funzione di attivazione è l'unità lineare rettificata (ReLU) che ha la caratteristica di prendere un input con valore reale negativo e lo fissa a zero (in altri termini sostituisce i valori negativi con zero).

Una rete neurale può essere costituita da tre tipi di nodi Figura 6):

1. i nodi di input forniscono informazioni dal mondo esterno alla rete. In questa fase non viene eseguito alcun calcolo in quanto si passano semplicemente le informazioni ai nodi nascosti;
2. i nodi nascosti non hanno alcuna connessione diretta con il mondo esterno (da qui il nome "nascosto"), eseguono calcoli e trasferiscono le informazioni dai nodi di input ai nodi di output. Una raccolta di nodi nascosti forma un "livello nascosto" e le reti tipicamente hanno decine o centinaia di livelli nascosti;
3. i nodi di output vengono definiti collettivamente "livello di output" e sono responsabili dei calcoli e del trasferimento delle informazioni dalla rete al mondo esterno.

Figura 6 Rete neurale con uno strato nascosto, dove l'output del neurone evidenziato è f(w1 × X1+w2 × X2+b)



In questo tipo di rete neurale si risolve un problema di classificazione binaria in cui un percettore multistrato può imparare dagli esempi forniti (dati di addestramento) e fare una previsione informata dato un nuovo punto dati.

Il processo di addestramento di una rete neurale multistrato, come già accennato è basato su un algoritmo di retropropagazione.

Lo schema di funzionamento della rete neurale, esemplificato a partire dalla Figura 6 fino alla Figura 9, rappresenta un ambiente di formazione supervisionato, il che significa che apprende dai dati di addestramento etichettati e impara dagli errori. Il compito del supervisore è quello di correggere la rete ogni volta che commette errori. Come già detto, nell'apprendimento supervisionato, il set di allenamento è etichettato, ciò significa che, per alcuni input, si conosce l'output desiderato/previsto (etichetta) e dopo che sono stati inizialmente attribuiti dei pesi causali (Figura 7) viene confrontato l'output con quello atteso (Figura 8) e l'errore è propagato al livello precedente ricalcolando i pesi fino all'eliminazione dell'errore (Figura 9).

Figura 7 Output non corretto dove il nodo V assume i pesi w1, w2 e w3

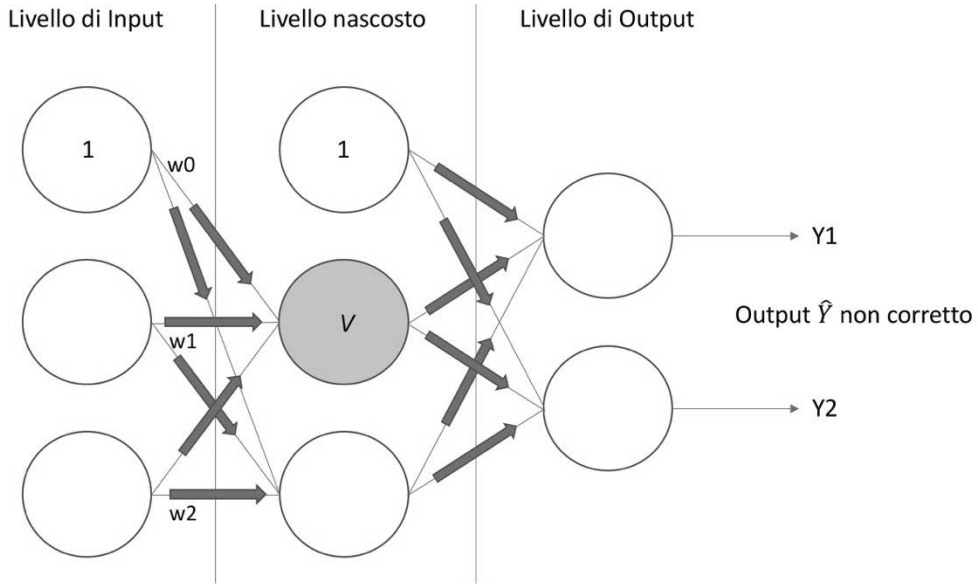


Figura 8 Schema di retropropagazione dell'errore e riattribuzione dei nuovi pesi w4, w5 e w6

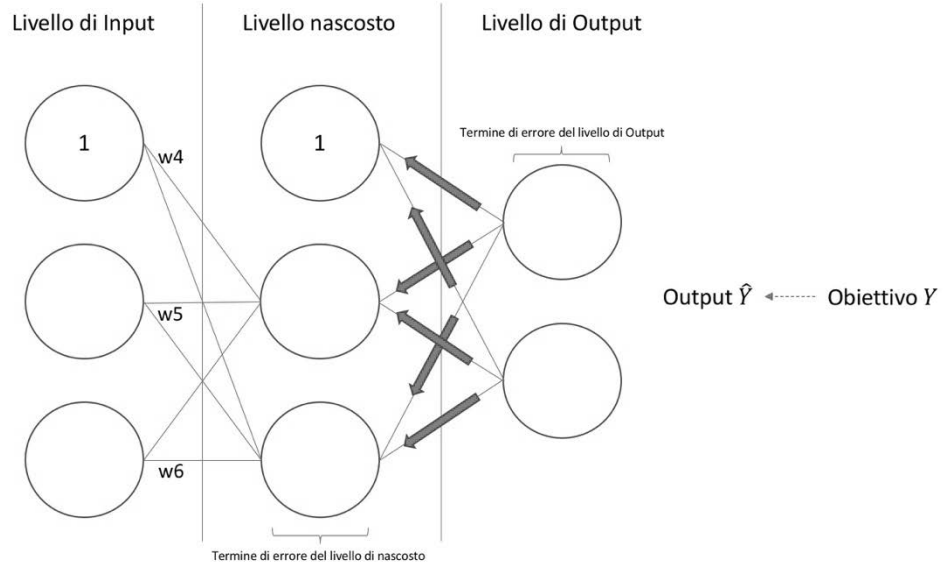
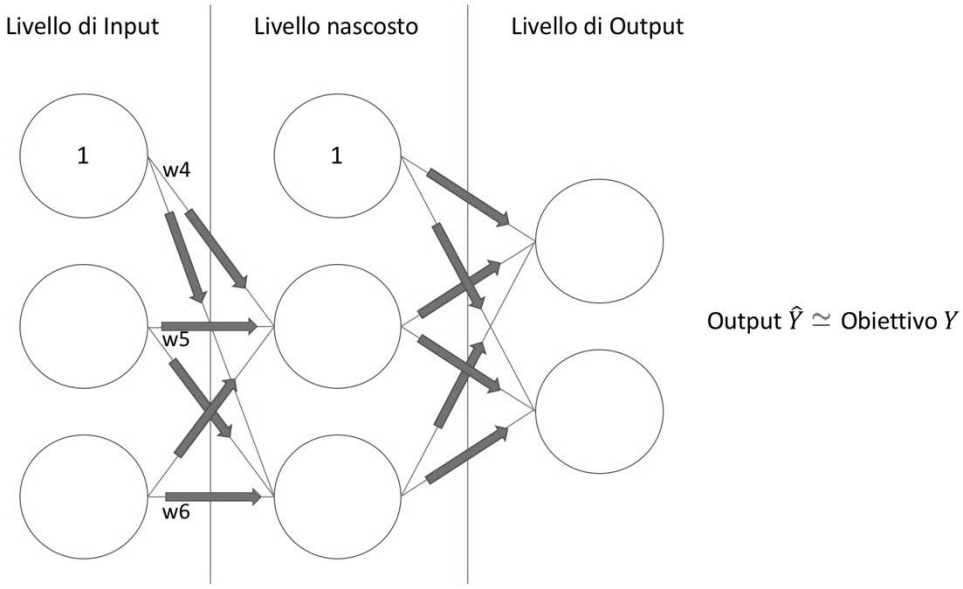


Figura 9 Output corretto in cui il valore di output $\hat{Y}$ è vicino al valore obiettivo Y

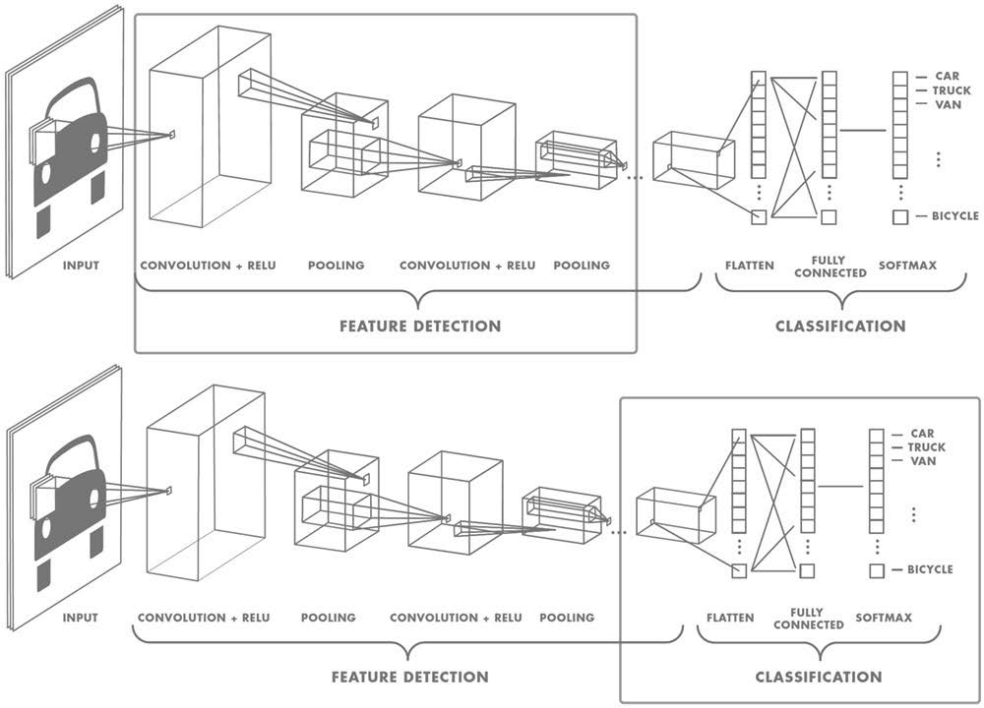


Una volta terminato il processo di addestramento descritto sopra, si è ottenuta una rete neurale "esperta" da considerarsi pronta per essere applicata a casi generali.

Attualmente la tipologia di rete neurale più popolare per l'apprendimento approfondito di immagini e video è la cosiddetta rete neurale convoluzionale (Krizhevsky, Sutskever and Hinton, 2012). Come le altre reti neurali, anche la convoluzionale è composta da un livello di entrata, uno di uscita e molti livelli nascosti nel mezzo.

Il processo di funzionamento delle reti neurali convoluzionali è distinto in due blocchi: estrazione delle caratteristiche dell'immagine e classificazione delle caratteristiche (Figura 10, fonte: Mathworks).

Figura 10 Processo delle reti neurali convoluzionali



La prima parte del processo ha come scopo l'estrazione delle caratteristiche dell'immagine eseguendo tre tipi di azione: la convoluzione (convolution) inserisce le immagini di input in una serie di filtri, ciascuno dei quali attiva determinate funzionalità dalle immagini; il raggruppamento (pooling), semplifica l'output attraverso la riduzione del numero dei parametri che la rete deve conoscere; l'unità lineare rettificata (ReLU) consente una formazione più rapida ed efficace rimappando i valori.

Queste tre operazioni vengono ripetute su decine o centinaia di livelli, con ogni livello che impara a rilevare caratteristiche diverse.

Una volta rilevate le caratteristiche dell'immagine, la rete neurale convoluzionale passa alla seconda fase che mira a classificare le caratteristiche estratte.

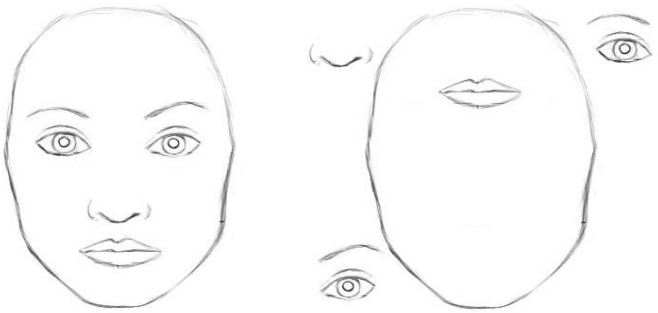
Il penultimo strato è un livello completamente connesso che genera un vettore con dimensione pari al numero di classi che la rete sarà in grado di prevedere. Questo vettore contiene la probabilità per ogni classe di qualsiasi immagine che viene classificata.

Il livello finale dell'architettura delle reti neurali convoluzionali utilizza una funzione di classificazione detta softmax per fornire l'output.

Le reti neurali e soprattutto le reti neurali convoluzionali sono estremamanete potenti e arrivano a fare cose che i computer non sono stati in grado di realizzare fino a pochi anni fa. Tuttavia, anch'esse presentano dei limiti fondamentali.

Ad esempio, la rappresentazione interna dei dati di una rete neurale convoluzionale non tiene conto di importanti gerarchie spaziali tra oggetti semplici e complessi, ciò rappresenta un limite perché sia nel riconoscimento facciale che, ad esempio, nel riconoscimento di oggetti non solo sono sufficienti le componenti per identificare una struttura, ma anche come questi sono spaziati e orientati l'uno rispetto all'altro.

Figura 11 Limiti delle reti neurali convoluzionali



Si prendano come esempio i disegni stilizzati di Figura 11, la rete neurale convoluzionale rileva dal volto di sinistra le componenti che lo compongono, l'ovale, due occhi, un naso e una bocca. La presenza di questi elementi per questo tipo di rete è un indicatore sufficiente per constatare che l'immagine raffigura una faccia. Purtroppo, per la rete anche il disegno di destra rappresenta un volto poiché compaiono tutti gli elementi necessari anche se non legati da relazioni spaziali orientazionali (v. sitografia).

Un altro svantaggio, molto spesso ignorato, delle reti neurali è il fatto che esse risentono in maniera marcata della presenza di valori anomali (i cosiddetti outliers). La presenza di outliers fa sì che la fase di apprendimento produca un sistema di pesi ($w1$, $w2$ …) che non riflettono le caratteristiche generali della maggior parte delle unità ma sono influenzati dalla presenza di caratteristiche atipiche. Nella fase di apprendimento è necessario quindi introdurre metodologie robuste che evitino questo problema (ad esempio Atkinson and Riani, 2000).

2.3.2 Metodi statistici di classificazione a massimo margine

Per la fase di classificazione, in alternativa alle rete neurali, possono essere usate le macchine a vettori di supporto (SVM, *Support Vector Machines*), o macchine kernel che sono delle metodologie di apprendimento supervisionato per la classificazione di pattern. Appartengono alla famiglia dei classificatori lineari generalizzati e sono noti anche come classificatori a massimo margine, poiché allo stesso tempo minimizzano l'errore empirico di classificazione e massimizzano il margine geometrico.

La caratteristica principale delle SVM è data dal fatto che esse, basandosi su semplici idee, permettono di giungere ad elevate prestazioni nelle applicazioni pratiche: sono, cioè, abbastanza semplici da analizzare matematicamente ma consentono di manipolare problemi molto complessi. Grazie a questa combinazione sono diventate popolari e diffuse in tempi molto rapidi.

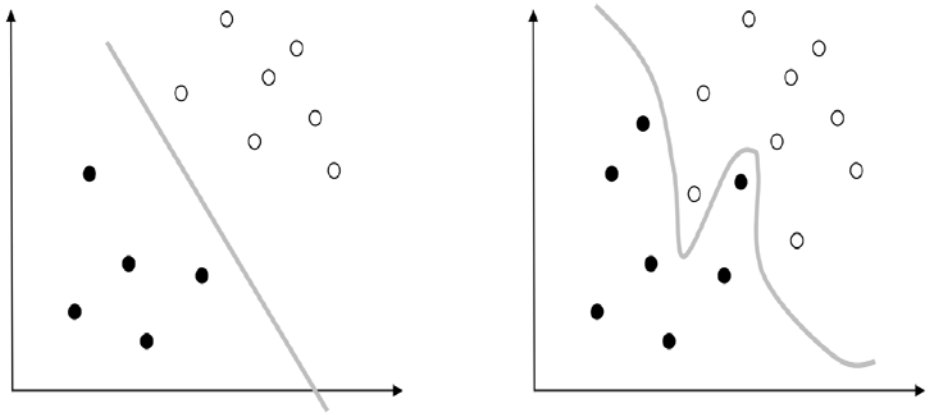
Le SVM hanno applicazioni in ambiti diversissimi. I più diffusi sono, ad oggi, il riconoscimento di strutture, la catalogazione di testi e l'identificazione di volti in immagini.

Dopo la fase di apprendimento, basata su un algoritmo che si può ricondurre ad un problema di programmazione quadratica con vincoli lineari, si ottengono i parametri dell'iperpiano ottimo di separazione e le SVM procedono alla classificazione vera e propria di nuovi dati. Questa fase viene chiamata fase di test e consiste nel collocare nella giusta classe un campione arbitrario, anche non appartenente ai dati di addestramento, proprio come farebbe la mente umana dopo aver immagazzinato informazioni tramite l'esperienza.

A differenza delle reti neurali che nel caso non lineare risultano complicate da addestrare, l'algoritmo delle SVM, con l'utilizzo degli spazi immagine e dei kernel, risulta essere molto efficiente e, in fase di ottimizzazione, non presenta problemi di minimi locali.

Si consideri quanto illustrato (Figura 12) dove sono presenti due classi (una in bianco e l'altra in nero) dove è possibile identificare la perfetta separabilità lineare e la perfetta separabilità non-lineare. Le SVM permettono di costruire "curve" ottimali, e risolvere problemi non lineari apprendendo dai dati.

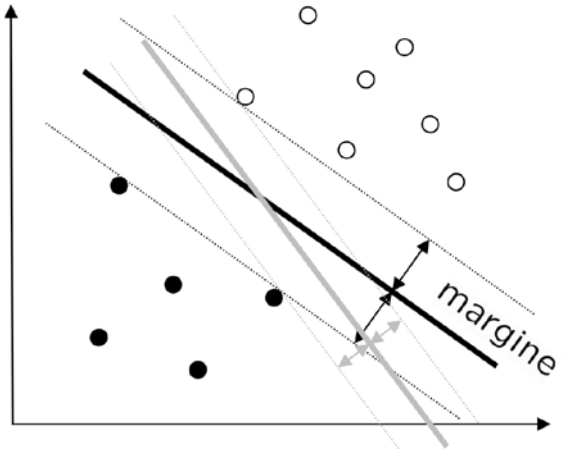
Figura 12 Separabilità lineare e separabilità non lineare



Le SVM sono nate per risolvere problemi di classificazione (supervisionata) rendendo flessibile il concetto di classificazione del massimo margine che viene illustrato, per semplicità, in un grafico linearmente separabile.

Il massimo margine identifica la retta in cui la distanza tra le due classi è la più grande possibile. Calcolarlo e studiarne le proprietà è compito relativamente semplice, ma omesso per non compromettere la lettura.

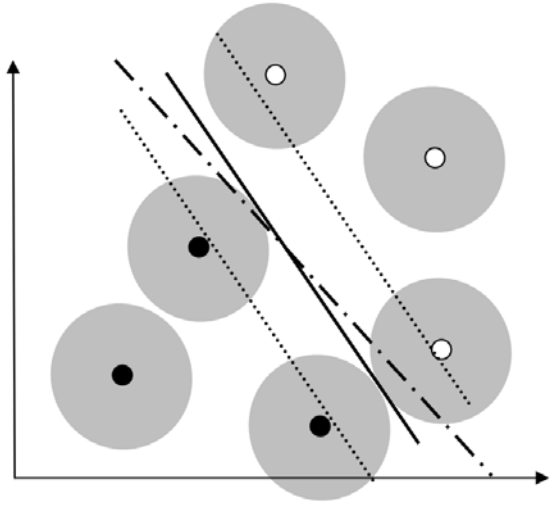
Figura 13 Concetto di massimo margine in due dimensioni



Il classificatore del massimo margine (Figura 13) sebbene sia elegante e semplice purtroppo non può essere applicato alla maggior parte dei dati pratici,

poiché richiede che le classi siano separabili perfettamente da un iperpiano lineare (che diventa una retta nel caso di 2 dimensioni).

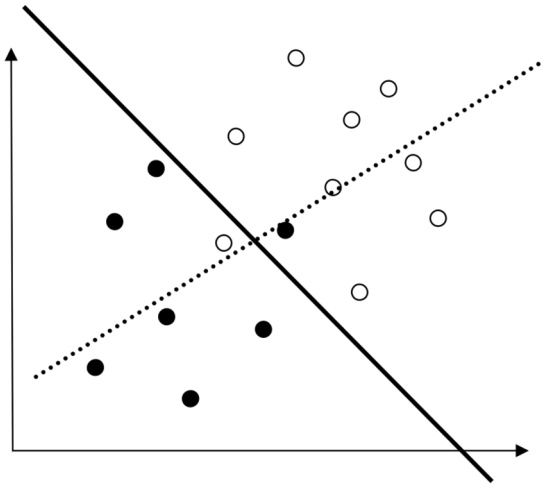
Figura 14 Classificatore di massimo margine in presenza di rumore



Il classificatore di massimo margine è ancora ottimale quando, per esempio, si è in presenza di "rumore", cioè di possibili perturbazioni dei dati su cui si effettua l'apprendimento. Nel contesto riportato in (Figura 14) l'iperpiano di massimo margine massimizza lo "spessore" della striscia rappresentata in figura. Il classificatore ottimo senza rumore sarebbe quello indicato con "tratto e punto", mentre se si considera il rumore abbiamo i "limiti" indicati con il tratteggio e la retta di supporto è indicata con tratto continuo. Quindi, in sintesi, la linea con tratto continuo rappresenta il vettore che individua l'iperpiano di massimo margine. Quest'ultimo, matematicamente, si trova come combinazione lineare dei vettori di supporto (le linee tratteggiate più esterne).

Il classificatore del massimo margine è stato migliorato dal classificatore con supporto di vettore quando si è in presenza di non perfetta separabilità lineare (Figura 15).

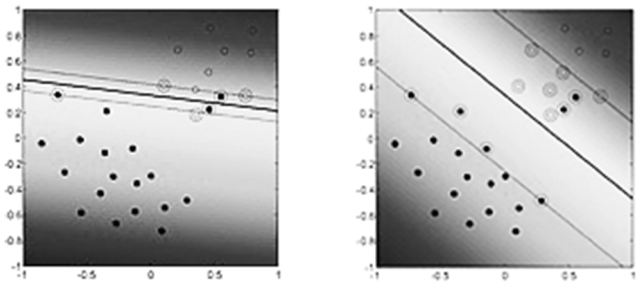
Figura 15 Esempio di non perfetta separabilità lineare



Se il data set di training non è linearmente separabile, nessun iperpiano potrà classificare correttamente tutti i punti, ma è chiaro che vi sono iperpiani migliori di altri.

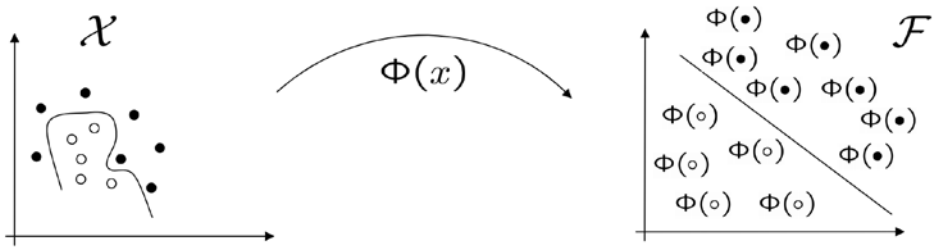
L'idea di fondo, in questi casi, è quella di minimizzare il numero di errori di classificazione e simultaneamente massimizzare il margine per i punti correttamente classificati. Questo avviene rendendo i margini "flessibili o soft" ed impostando un limite di tolleranza "variabile". Tuttavia, siccome il problema può coinvolgere un numero di combinazioni che è potenzialmente grandissimo occorre introdurre un vincolo di regolarizzazione. Questo modo di procedere ammette, quindi, che vi siano punti nel dataset che saranno classificati in modo errato, come riportato in Figura 16.

Figura 16 Classificazioni con vincolo di tolleranza



Il grado di tolleranza all'errore di classificazione viene controllato dal fattore di regolarizzazione. Tanto maggiore sarà la regolarizzazione, tanto peggiore sarà la classificazione nel dataset di addestramento, ma tanto minore sarà il rischio di *overfitting*[23]. In molti casi anche il soft margin è inadeguato poiché il legame fra i dati è non lineare e dunque nessuna legge lineare può "catturare" la regolarità dei dati, come in Figura 17.

Figura 17 Inadeguatezza del soft margin per dati non lineari



Con un'opportuna trasformazione è possibile però rendere i dati linearmente separabili. Nella Figura 17 χ è lo spazio degli attributi di ingresso e $\mathcal{F}$ è quello delle caratteristiche.

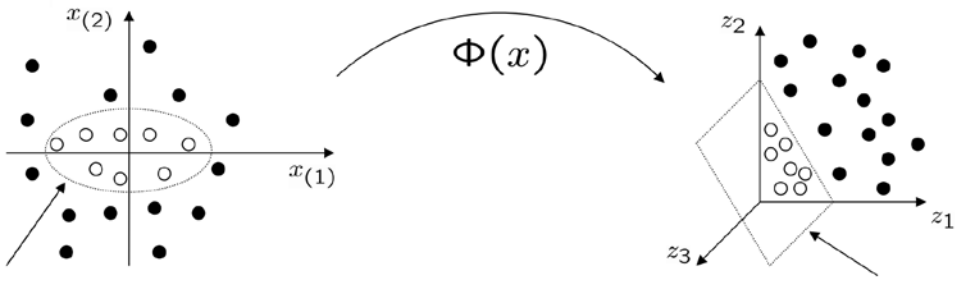
Allora si può pensare ad un algoritmo in due passi:

1. trasformazione non lineare (*feature mapping*) tramite la $\Phi(x)$: $\chi \rightarrow \mathcal{F}$;
2. classificazione lineare nello spazio delle caratteristiche.

A volte per trovare un classificatore lineare occorre anche aumentare la dimensione del problema, come illustrato di seguito passando da uno spazio degli attributi di dimensione 2 ad uno spazio delle caratteristiche di dimensione 3, ma linearmente separabile da iperpiano, in altre parole l'ellisse in due dimensioni viene trasformato in un piano a 3 dimensioni (Figura 18).

[23] In statistica e in informatica, si parla di overfitting (in italiano: eccessivo adattamento) quando un modello statistico molto complesso si adatta ai dati osservati (il campione) perché ha un numero eccessivo di parametri rispetto al numero di osservazioni.

Figura 18 Esempio di ampliamento delle dimensioni spaziali



Le SVM vere e proprie effettuano le trasformazioni descritte quando il numero di dimensioni è molto elevato. Per effettuare le trasformazioni si utilizzano i cosiddetti *kernel* di cui i più noti sono quelli: polinomiali (omogenei e non omogenei), Gaussiani, sigmoidale.

Occorre la seguente considerazione finale: una SVM con "kernel gaussiano" (Vapnik, 1998) è equivalente a certo tipo di rete neurale (base radiale), con funzione di attivazione gaussiana. Tuttavia il numero effettivo di neuroni (che sono i SV) è automaticamente scelto dall'algoritmo e la localizzazione dei "centri" è automaticamente scelta dall'algoritmo.

2.3.3 Metodi statistici di classificazione robusta

Un terzo approccio al riconoscimento facciale prevede, per la fase di classificazione, l'utilizzo di metodi di natura ancora più strettamente statistica.

Il principio di questo approccio consiste nel tagliare la rete neurale convoluzionale all'altezza dell'ultimo livello di estrazione delle caratteristiche del volto e applicare a questi risultati parziali un mix di metodologie di statistica robusta sfruttando criteri di analisi discriminante.

L'analisi discriminante è una disciplina statistica che consente di raggruppare oggetti o osservazioni in classi distinte e per allocare le nuove osservazioni nelle classi precedentemente definite. Quando la classificazione delle osservazioni è riferita al machine learning si parla di riconoscimento di pattern.

Le tecniche di analisi discriminante si suddividono in supervisionate e non supervisionate. Nell'ambito del riconoscimento facciale, le tecniche che risultano più adeguate sono quelle non supervisionate, più nel dettaglio la tecnica di classificazione di cluster analysis robusta TCLUST (García-Escudero, Gordaliza, Matrán and Mayo-Iscar, 2008).

Quando si parla di cluster analysis in termini generali si fa riferimento ad una famiglia di algoritmi molto estesa che tipicamente danno origine a risultati molto spesso fortemente influenzati dal metodo scelto e le prestazioni di ciascun

metodo dipendono dal modello probabilistico sottostante e dalla quantità di osservazioni che si allontanano da tali distribuzioni sottostanti.

Il modello che più si presta alla classificazione delle caratteristiche estratte dai volti è il TCLUST poiché sì è dimostrato un algoritmo resistente al variare delle distribuzioni sottostanti e capace di gestire tutte le caratteristiche estratte anomale che potrebbero portare all'errato riconoscimento dei volti (barba, baffi, occhiali, ecc…).

Più formalmente, dato un insieme *n* di numero di unità statistiche (ad esempio facce) su cui sono stati rilevati i valori di *p* variabili (ad esempio *p* tratti somatici), e fissato un predeterminato numero di gruppi pari a *k*, l'algoritmo TCLUST ha una funzione target (Equazione 2-3) caratterizzata dal taglio delle osservazioni anomale dato da $n(1-\alpha)$ dove α è la frazione dei valori anomali e da un sistema vincoli sugli autovalori della matrici di covarianze indicati con $\lambda_l(\Sigma_j)$ con $j = 1, \dots, k$ e $l = 1, \dots, p$ (Equazione 2-1 e Equazione 2-2). Il vincolo imposto da TCLUST sul rapporto tra il massimo autovalore (M_n) ed il minimo autovalore (m_n) in modo tale che $M_n/m_n \leq c$, con *c* fissato a priori, consente di evitare di incappare in soluzioni spurie.

Equazione 2-1

$$M_n = \max_{j=1,\dots,k} \max_{l=1,\dots,p} \lambda_l(\Sigma_j)$$

Equazione 2-2

$$m_n = \min_{j=1,\dots,k} \min_{l=1,\dots,p} \lambda_l(\Sigma_j)$$

Equazione 2-3

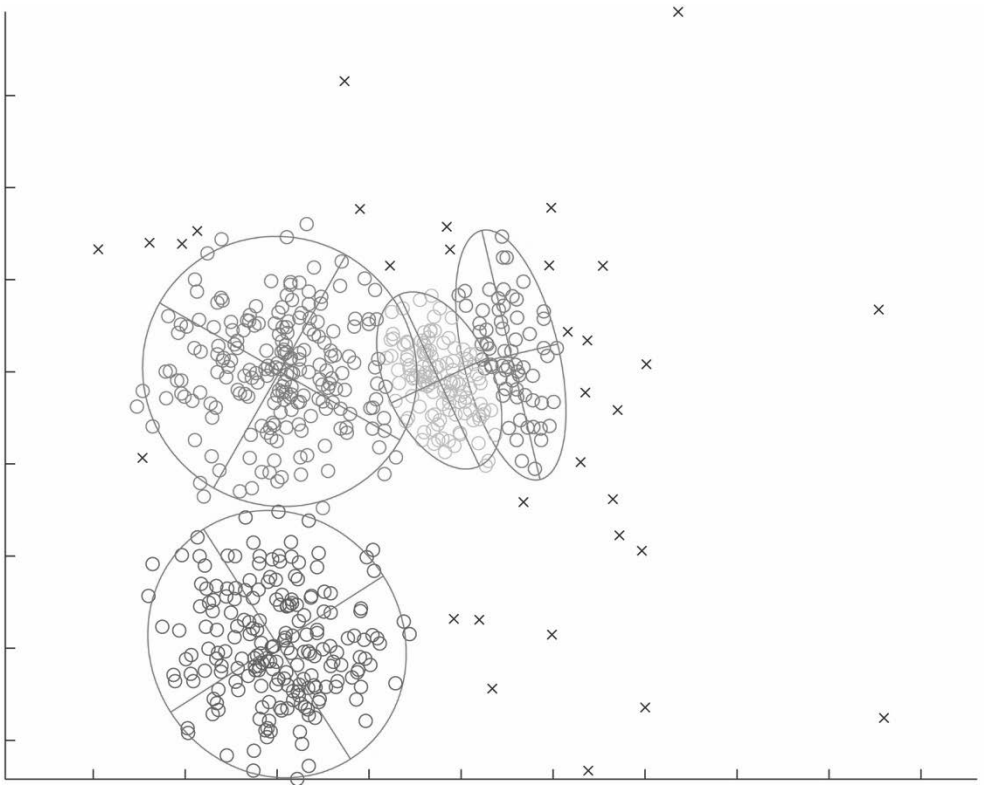
$$\sum_{j=1}^{k} \sum_{i \in R_j} \log(\pi_j f(x_i; \mu_j, \Sigma_j))$$

Dove $R_j = n(1-\alpha)$, e μ_je Σ_j rappresentano rispettivamente il centroide e la matrice di covarianze del gruppo *j* .

Con tale procedura si ottiene una classificazione delle osservazioni in gruppi distinti che tiene conto di eventuali valori anomali e che massimizza la somiglianza delle osservazioni all'interno di ciascun gruppo e minimizza le somiglianze delle osservazioni che appartengono a gruppi differenti. In Figura 19 si può osservare un esempio di raggruppamento di 630 immagini appartenenti a quattro individui. Ciascun circoletto rappresenta la sintesi delle caratteristiche di ciascuna immagine catturata ed elaborata dal primo gruppo di strati della rete

neurale, i cerchi e le ellissi più grandi sono i raggruppamenti delle immagini associate alla stessa persona. Le immagini che mostrano caratteristiche estratte non conformi o anomale vengono indicate con il simbolo "×" e rimosse dalla classificazione. La rimozione delle osservazioni anomale è la caratteristica distintiva della statistica robusta che permette alle metodologie utilizzate, riviste in chiave robusta, di non essere contaminate da osservazioni che si presentano molto diverse da quelle attese o per errori di rilevazioni o per natura differente.

Figura 19 Classificazione delle immagini dei volti appartenenti a quattro soggetti distinti (rappresentati dalle ellissi) mostrata in due dimensioni.



L'applicazione del TCLUST nel riconoscimento facciale si basa su di un processo iterativo che alloca le nuove osservazioni ai gruppi (cluster) più appropriati. In altre parole, le caratteristiche estratte dalle immagini riferite allo stesso volto vengono raggruppate in un singolo cluster. Esisteranno, quindi, un numero di cluster pari al numero dei soggetti su cui è stato fatto l'addestramento. Una volta ottenuti i gruppi, ogni nuovo volto viene confrontato con le caratteristiche contenute in ciascun gruppo, in caso positivo la nuova immagine

viene inserita nel gruppo di immagini appartenente all'individuo catturato in foto, in caso negativo il soggetto viene riconosciuto come nuovo soggetto e va a creare un nuovo gruppo che allargherà la base di confronto delle successive iterazioni.

Data l'onerosità computazionale di tale tecnica di clustering è necessario un pretrattamento dei dati attraverso un modello statistico di riduzione delle dimensioni. Per fare ciò viene utilizzata l'analisi delle componenti principali (ACP) robusta. L'ACP è una tecnica standard utilizzata per approssimare i dati originali con vettori di caratteristiche dimensionali inferiori. La procedura matematica delle componenti principali mira a trasformare il numero delle variabili potenzialmente correlate in un numero più piccolo di variabili incorrelate con lo scopo di non perdere informazioni e contestualmente di ridurre la mole dei dati.

L'approccio di base delle componenti principali può essere pensato come un ellissoide *n*-dimensionale che si adatta ai dati originari, in cui ciascun asse rappresenta una componente principale. Gli assi più piccoli dell'ellissoide implicano una varianza molto piccola lungo l'asse stesso il che comporta una minima perdita di informazione nel caso in cui si decida di ometterlo.

Il procedimento robusto per trovare l'ellissoide si articola in più fasi:

1. sottrazione della mediana di ogni variabile dal set di dati al fine di centrare i dati attorno al centro della nuvola dei punti;
2. calcolo della matrice di covarianza robusta dei dati e dei suoi autovalori e autovettori;
3. normalizzazione di ciascuno degli autovettori ortogonali per trasformarli in vettori unitari.

Una volta eseguite queste operazioni, ciascuno degli autovettori reciprocamente ortogonali può essere interpretato come un asse dell'ellissoide adattato ai dati. In questo modo si ottiene la nostra matrice di covarianza in una forma diagonalizzata con gli elementi diagonali che rappresentano la varianza di ciascun asse. La proporzione della varianza rappresentata da ciascun autovettore può essere calcolata dividendo l'autovalore corrispondente con tale autovettore per la somma di tutti gli autovalori.

Attraverso l'individuazione delle componenti principali il TCLUST può operare su un set di dati sensibilmente più piccolo seppur ancora altamente informativo e quindi velocizzare sensibilmente il processo di classificazione.

2.4 La struttura software

Lo sviluppo di un sistema di riconoscimento facciale necessita di diversi ambienti software sia per il riconoscimento facciale in senso stretto sia per l'archiviazione dei dati raccolti.

Nell'ambito del progetto, per la programmazione delle reti neurali, lo sviluppo dei modelli statistici utilizzati e la gestioni dei dati sono stati utilizzati gli ambienti: Matlab, Python e MySQL.

La scelta di sviluppare un processo di riconoscimento facciale è ricaduta su questi tre ambienti per la loro efficienza e per l'esperienza maturata nel corso degli anni dagli sviluppatori in questi ambiti.

2.4.1 L'ambiente MATLAB

MATLAB è un linguaggio di alto livello dotato di un ambiente interattivo per il calcolo scientifico, la visualizzazione e la programmazione numerica. MATLAB permette di analizzare dati, sviluppare algoritmi e creare modelli e applicazioni. Gli usi tipici includono:

- matematica e computazione;
- sviluppo di algoritmi;
- modellazione, simulazione e prototipazione;
- analisi dei dati, esplorazione e visualizzazione;
- grafica scientifica e tecnica;
- sviluppo di applicazioni, compresa la costruzione di interfacce grafiche utente.

Negli ambienti accademici rappresenta uno strumento didattico largamente diffuso per corsi introduttivi e avanzati in matematica, statistica e ingegneria.

Oltre alle funzioni native MATLAB ha una ampia dotazione di pacchetti aggiuntivi detti Toolbox che consentono di utilizzare algoritmi fortemente specializzati per risolvere particolari classi di problemi.

L'ambiente MATLAB è composto da cinque parti principali:

- un linguaggio matrice/matrice di alto livello con istruzioni di flusso di controllo, funzioni, strutture dati, input/output e funzioni di programmazione orientate agli oggetti;
- l'ambiente di lavoro, definito come l'insieme degli strumenti e dei servizi a disposizione dell'utente, ad esempio le funzionalità per la gestione delle variabili nel proprio spazio di lavoro, l'importazione e l'esportazione di dati e il debugging;

- gestore grafico per la rappresentazione di dati bidimensionali e tridimensionali, per l'elaborazione di immagini e per la creazione di interfacce utente;
- la libreria delle funzioni matematiche e la libreria che permette di scrivere programmi C e Fortran che interagiscono con MATLAB.

Grazie al suo ambiente di programmazione flessibile e alla sua capacità di interfacciarsi con i database e altri linguaggi di programmazione, MATLAB è la scelta più vantaggiosa per gestire la complessità dello sviluppo di un software per il riconoscimento facciale. Inoltre la necessità di trattare i dati con metodologie robuste è avvenuta tramite l'utilizzo di uno dei toolbox più importanti in questo ambito statistico, il Flexible Statistics Data Analysis (FSDA) MATLAB Toolbox™, sviluppato dal Centro Interdipartimentale di Ricerca di Statistica Robusta (Ro.S.A.) dell'Università di Parma (http://rosa.unipr.it) e dal Joint Research Centre (JRC) della Commissione Europea di Ispra. Questo toolbox ha ottenuto il premio come miglior contributo italiano durante il MATLAB Expo del 2016[24].

2.4.2 L'ambiente Python

Python è un linguaggio di programmazione interpretato, orientato agli oggetti e di alto livello con semantica dinamica. Python, grazie alla sua capacità di ottimizzare il modo in cui un processo di elaborazione gestisce le informazioni in memoria combinata con la digitazione dinamica e il binding dinamico, lo rendono adatto sia al cosiddetto sviluppo rapido delle applicazioni, sia al suo utilizzo come linguaggio di scripting.

Python è dunque un linguaggio di programmazione nella sua essenza molto snello che ha però la possibilità di essere arricchito attraverso numerose librerie dedicate ad una ampia fattispecie di problemi.

Da notare che Pyhton, grazie proprio alla grande disponibilità di librerie e alla portabilità cross-platform, è stato scelto come linguaggio ufficiale per l'implementazione e lo sviluppo del software su dispositivi Raspberry PI (v. sezione 2.5).

[24] Il toolbox FSDA è scaricabile gratuitamente dall'indirizzo web http://rosa.unipr.it/fsdadownload.html

2.4.3 Il database MySQL

Il processo di riconoscimento facciale, come già accennato, avviene attraverso il confronto tra una nuova immagine acquisita e una immagine già esistente. Il contenitore che archivia e gestisce tutte le immagini esistenti e accetta quelle nuove è detto database. Un database è una raccolta di informazioni che è organizzata in modo che possa essere facilmente accessibile, gestita e aggiornata.

I dati sono organizzati in righe e colonne che vanno a formare le tabelle e sono indicizzati per facilitare la ricerca di informazioni rilevanti. I dati vengono aggiornati, espansi ed eliminati quando vengono aggiunte nuove informazioni.

Tipicamente i database sono di tipo relazionale e sono costituiti da un insieme di tabelle. Le tabelle sono legate tra loro da una o più relazioni che permettono di archiviare più efficientemente i dati e richiamarli in modo più efficace.

Il linguaggio utilizzato per la progettazione dei database relazionali è lo Structured Query Language (SQL) e permette di:

- creare e modificare gli schemi del database;
- inserire, gestire e modificare dati;
- interrogare i dati archiviati;
- creare e gestire strumenti di controllo ed accesso ai dati.

Il grande vantaggio che si ha nell'utilizzo dei database relazionali è la possibilità di espanderne la struttura in qualsiasi momento. Anche dopo la creazione del database originale, infatti, è possibile aggiungere una nuova categoria di dati senza richiedere la modifica di tutte le applicazioni esistenti.

Il database più popolare del mondo, come sottolineato dal produttore, è MySQL. I motivi per cui MySQL ha avuto una tale diffusione sono da individuare nei seguenti fattori:

- facilità di utilizzo;
- un solido sistema di privilegi utente e password criptate che lo rendono sicuro sul fronte delle potenziali intrusioni esterne;
- open source, quindi gratuito;
- velocità operative;
- possibilità teorica di archiviazione pari a 8 TB di dati;
- ottima gestione della memoria.

MySQL non ha solamente vantaggi ma anche svantaggi, quali una serie di ridotte funzionalità rispetto ad altri database di natura commerciale. Le sue funzionalità tuttavia, anche se ridotte, sono ampiamente sufficienti per la maggior parte delle esigenze, comprese quelle legate alla gestione dei dati di un processo riconoscimento facciale.

2.5 Lo sviluppo hardware

Fino a questo punto si è parlato di quali sono le architetture metodologiche e software per avviare un processo di riconoscimento facciale, ma non si è ancora fatto accenno a cosa è necessario dal punto di vista hardware.

La componente hardware nel campo del riconoscimento facciale merita un approfondimento poiché non è di natura banale. Per molto tempo il riconoscimento facciale in tempo reale non ha potuto avere applicazioni al di fuori degli ambienti di ricerca a causa della straordinaria capacità di calcolo richiesta.

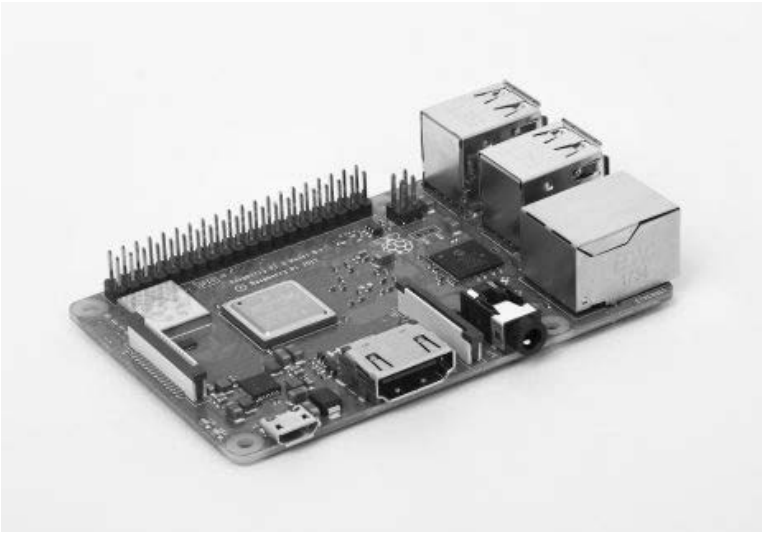
Solamente negli ultimi anni, con la nascita di nuove architetture di calcolo, è stata resa possibile la sperimentazione del riconoscimento facciale nel mondo reale riducendo i tempi di elaborazione del processo da ore o talvolta giorni in pochi decimi di secondo. In questo contesto, verrà descritto l'architettura hardware per la cattura e il trasferimento delle immagini e per l'elaborazione dei fotogrammi.

2.5.4 Gli strumenti di cattura delle immagini

Appare ragionevole pensare che un sistema di riconoscimento facciale debba essere in grado di monitorare interi ambienti di ampia superficie e partendo da questo presupposto è altrettanto naturale pensare che non è possibile installare un elaboratore elettronico in prossimità di ciascuna area da tenere sotto controllo. Per tale motivo è necessario individuare un hardware che non solo catturi le immagini, ma che sia anche in grado di eseguirne una pre-elaborazione e di inviare un risultato grezzo ad una unità di calcolo centralizzata.

Lo strumento utilizzato per questa fase è il Rasperry PI 3 Model B+ (Figura 20) dotato di una fotocamera ad infrarossi. Il Raspberry PI è un single-board computer di ridotte dimensioni (65mm × 30mm × 5mm) dotato di un processore ARMv8 quad-core da 1.4GHz a 64-bit, 1GB di ram, connettività wireless, LAN e Bluetooth con la possibilità di ospitare una memoria microSD fino a 32GB, il tutto progettato per ospitare sistemi operativi basati sul kernel Linux o RISC OS.

Figura 20 Raspberry PI 3 Model B+



La cattura delle immagini è affidata ad una interfaccia Quimat (Figura 21) dotata di:

- sensore da 5 megapixel con risoluzione fino a 1080p;
- campo visivo da 76 gradi per fornire ampie inquadrature dell'ambiente monitorato;
- massima apertura f/1.8 che assicura la qualità delle immagini anche in caso di bassa illuminazione;
- filtro IR-CUT per evitare le distorsioni cromatiche della luce del giorno;
- una coppia di LED ad infrarossi per illuminare i soggetti in caso di scarsa luce ambientale.

Come sarà mostrato nel prossimo capitolo in modo più esaustivo, il binomio hardware descritto garantisce l'acquisizione continua di immagini dell'ambiente scattando foto con ad alta frequenza e inviandole attraverso una connessione di rete protetta ad una unità di calcolo centrale.

Figura 21 Quimat Camera per Raspberry Pi

2.5.5 L'unità di calcolo centrale

Le elaborazioni per il riconoscimento facciale passano attraverso algoritmi di calcolo molto dispendiosi dal punto di vista computazionale e da ciò consegue la necessità di utilizzare macchine molto potenti. Naturalmente la potenza della macchina deve essere relazionata al numero di soggetti che frequentano l'ambiente sotto osservazione e alla quantità dei soggetti che sono censiti all'interno del database. Prescindendo dalle proporzioni del progetto, l'innovazione tecnologica che ha permesso di poter eseguire le identificazioni in tempo reale, non è stato l'incremento della potenza di calcolo dei processori (CPU), ma l'introduzione di un nuovo approccio al calcolo scientifico basato sullo sfruttamento della potenza di calcolo di alcuni tipi di schede grafiche.

Recentemente la casa costruttrice di schede grafiche NVIDIA ha introdotto su mercato una nuova architettura per il calcolo parallelo su processori grafici (GPU) chiamata CUDA (Compute Unified Device Architecture).

I linguaggi di programmazione disponibili nell'ambiente di sviluppo per CUDA, sono estensioni dei linguaggi più famosi come il C, Python, Java, Fortran e MATLAB. Grazie all'utilizzo di CUDA le GPU diventano architetture aperte come le CPU ma, a differenza di queste ultime, sono caratterizzate da una struttura di calcolo parallela basata su diversi core, ciascuno dei quali capace di eseguire centinaia di processi contemporaneamente.

La nuova architettura di elaborazione in parallelo su GPU, sta rivoluzionando il paradigma della centralizzazione del calcolo su CPU evolvendosi verso il

concetto del "co-processing" su CPU e GPU riducendo i tempi di calcoli di diversi ordini di grandezza e rendendo così utilizzabili metodi statistici fino a qualche anno fa destinati solo alle analisi a posteriori dei fenomeni. Per i dettagli relativi all'utilizzo di questa tecnologia nell'ambito del processo di riconoscimento facciale si rinvia al capitolo successivo.