Corso di NOzioni di Visione Artificiale – NOVA23

Lezione 1
Introduzione al Machine Learning

prof. Vito Roberto
Dipartimento di Matematica, Informatica e Fisica (DMIF)
Università di Udine
E-mail: vito.roberto@uniud.it

1.- Intelligenza Artificiale - Artificial Intelligence, AI

L'AI è la capacità dei computer di simulare azioni degli esseri viventi ritenute 'intelligenti', quali: comprendere il linguaggio naturale; risolvere problemi in base alle conoscenze acquisite; percepire e comprendere stimoli sensoriali; partecipare a giochi; pianificare il controllo di robot; imparare nozioni e azioni.

L'AI trae ispirazione dai sistemi viventi (hardware biologico) per emularne le prestazioni operando su supporti tecnologici (hardware digitale, HW). McCarthy definisce l'AI *"La scienza e le tecnologie volte alla realizzazione di macchine intelligenti, e in particolare programmi di computer intelligenti".*

Evidentemente si tratta di terminologia per dare significati che non sono univoci. Col termine AI si indica da un lato lo studio multidisciplinare della struttura di organi e comportamenti di esseri viventi; ciò avviene in collaborazione con la Neurofisiologia; la Psicologia; la Logica; la Statistica. Dall'altro si indica la progettazione di sistemi hw/sw che compiono azioni come si è illustrato in precedenza.

I progressi nell'IA hanno aperto nuove prospettive per il mondo produttivo e le amministrazioni. I computer trattano grandi quantità di dati (*big data);* compiono azioni ripetitive ben oltre le capacità degli esseri umani. Permettono di ridurre la manodopera poco qualificata, e spostare il lavoro umano verso compiti con maggiore apporto di conoscenze. Accade perciò che le prestazioni dei sistemi umani costituiscono un termine di confronto con le analoghe prestazioni dei computer.

Attualmente è divenuto essenziale l'apporto dell'IA alla nostra vita quotidiana; diamo solo alcuni esempi di aree d'interesse per i sistemi intelligenti:
- *Comprensione del linguaggio naturale:* effettuare traduzioni da una lingua a un'altra in contesti ben definiti; interloquire con persone, macchine in una lingua corrente; trasferire contenuti da un supporto mediale a un altro;…
- *L'IA Basata sulla conoscenza (Knowledge-based AI).* Due sotto-settori rilevanti sono la *Rappresentazione della conoscenza (Knowledge Representation)* e il *Ragionamento Automatico (Automated Reasoning).* La prima propone linguaggi formali per descrivere oggetti, concetti, relazioni e costruire la *Base di conoscenze (Knowledge base)* relativa a un dominio. Il Ragionamento automatico comprende tecniche per *estendere* le basi di conoscenze tramite meccanismi logici di deduzione, induzione, classificazione. Dal punto di vista applicativo, i *Sistemi a base di conoscenze (Knowledge-based Systems, KBS)* risolvono problemi complessi che richiedono la consultazione di archivi di dati e l'individuazione di soluzioni, come per il supporto a diagnosi mediche; o la consulenza personalizzata per servizi e tariffe;
- *La Percezione Artificiale.* Sotto-settori sono la *Visione Artificiale (Machine vision);* il *Riconoscimento del parlato (Speech understanding);* la *Fusione multisensoriale (Multisensor Data Fusion);*
- *I Giochi (Gaming).* Da decenni sono un terreno di confronto delle prestazioni dei sistemi intelligenti. Già nel 1997 il programma Deep Blue ha battuto il campione di scacchi Gary Kasparov;
- *La Robotica Intelligente*. Richiede la fusione multisensoriale e il controllo di componenti: motori, bracci meccanici,… Progressi rilevanti si sono realizzati per la guida autonoma di auto e i droni *(Unmanned Autonomous Vehicles, UAV);* i veicoli guidati a distanza *(Remotely Operated Vehicles, ROV)* per operare in ambienti ostili come il pianeta Marte;
- *L'Apprendimento Automatico (Machine Learning, ML).* E' trasversale ai settori dell'IA cui si è accennato. Ciò è in analogia con l'apprendimento che accompagna le funzioni cognitive negli esseri viventi, e ne migliora le prestazioni. Ce ne occuperemo con dettaglio nel seguito.

2.- Apprendimento Automatico (Machine Learning, ML)
Branca dell'IA con il compito di progettare e realizzare procedure che, sottoposte a stimoli esterni molteplici (segnali), rispondono migliorando progressivamente le loro prestazioni per risolvere un problema.
Le procedure di ML sono composte da algoritmi che lavorano con *dati di apprendimento (training set)*, e producono conclusioni *(decision)* senza che le stesse procedure siano state progettate esplicitamente per farlo. Accenniamo ad uno schema di procedura che svilupperemo nel seguito.

- A un computer sono presentati come stimoli *grandi quantità di dati digitalizzati (big data)* – registrazioni vocali; firme, testi manoscritti; documenti a stampa; immagini; sequenze video e audio-video; grafica manuale e a stampa;…
- Gli algoritmi di ML analizzano i dati; ne rilevano le caratteristiche; effettuano stime statistiche. Sulla base dei risultati decidono come classificare i dati in categorie e stimano l'errore di classificazione. *La valutazione dell'errore è l'elemento essenziale di una procedura di apprendimento;*

- Sulla base dell'errore stimato, la stessa procedura ricalcola i parametri del proprio funzionamento, e li modifica per ridurre l'errore;
- L'intero ciclo di calcolo a partire dalla presentazione degli stimoli viene ripetuto con i nuovi parametri, sino a quando l'errore diventa sufficientemente piccolo.

Gli schemi di ML sono stati applicati a problemi reali. Ad esempio, nel dominio della Visione Artificiale (Computer Vision) si è dimostrato più efficiente addestrare una macchina a compiere un'azione visiva, anziché richiedere che programmatori descrivessero ogni azione da compiere in ogni singola scena. Si può affermare in altri termini che una procedura di ML è in grado di *apprendere una funzione di decisione*, cioè una trasformazione *(map)* che mette in corrispondenza un insieme di stimoli in input con un insieme di risposte in output.
Vi sono molteplici *strategie di apprendimento:* ciascuna stabilisce una metodologia di lavoro per progettare e realizzare *modelli* di apprendimento. Questi sono *schemi astratti di strutture e procedure di calcolo*. Una struttura specifica i *componenti* del calcolo e la loro disposizione spaziale *(topologia);* le *connessioni* tra loro; le regole matematiche che stabiliscono i *flussi delle informazioni.* A loro volta le strutture danno origine ad *architetture di calcolo* quando si sono specificati i loro parametri numerici. Si giunge alle fasi di progettazione realizzativa della procedura, costituita da algoritmi e dati per svolgere determinate azioni. Le fasi successive riguardano l'esecuzione e test della procedura; richiedono una nuova metodologia di lavoro, che sarà presentata nel seguito.

2.1 Strategie di apprendimento

(a) Apprendimento supervisionato (Supervised learning). Nei casi in cui sia noto e disponibile a-priori un insieme di soluzioni valide, queste vengono etichettate *(labelled)* e inserite nell'insieme dei dati di apprendimento *(training set).* Al computer si presentano in input esempi di dati/segnali con le rispettive soluzioni valide. Si attiva la rete dall'input all'output e si valuta la differenza (errore) della risposta ottenuta rispetto alla risposta nota a-priori. Si applica una *regola di apprendimento (learning rule),* cioè si modificano i parametri dell'architettura di rete *(updating)* per ridurre l'errore. Si itera la procedura sino a raggiungere una risposta soddisfacente, cioè vicina a quella valida entro un margine di errore prefissato.
(b) Apprendimento non supervisionato (Unsupervised learning). Nei casi in cui non sia disponibile un insieme di soluzioni valide note a-priori si può ugualmente definire una strategia di apprendimento, in modo tale che sia la stessa procedura a scoprire regolarità e criteri di aggregazione dei dati di input. Si dice che in tali casi la rete si auto-organizza *(self-organizing map).* Scoprire regolarità può essere di per sé lo scopo dell'apprendimento; oppure può essere uno strumento per scoprire caratteristiche comuni dei segnali/immagini di input *(feature learning)* come accade nella Computer Vision.
(c) Apprendimento per conferma (Reinforcement learning). Un programma interagisce con un ambiente esterno dinamico, nel quale deve raggiungere un obiettivo. Esempi sono la guida automatica di un veicolo robotizzato che deve evitare ostacoli; oppure un gioco contro un avversario. Mentre naviga nello spazio virtuale, il programma acquisisce risposte *(feedback)* che agiscono come riscontri di validità: possono assegnare penalità *(penalty),* nei casi negativi che un algoritmo cercherà di minimizzare; o ricompense (*reward)* che cercherà di massimizzare.

3.- Modelli matematici dell'apprendimento

Adottare una strategia di apprendimento richiede un modello matematico per progettare le fasi di calcolo. Vari tipi di modelli sono stati proposti per il ML.

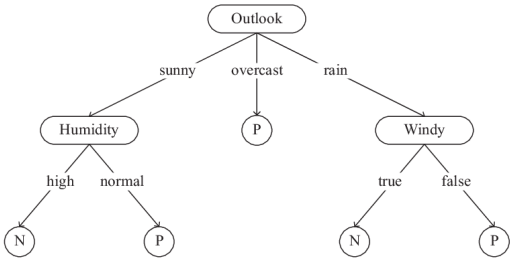
3.1 Reti Neurali Artificiali (ANN)

Una ANN è un gruppo di nodi tra loro connessi da legami *(link, edge)* in analogia con quanto si verifica per i neuroni del cervello. Nel modello ogni nodo rappresenta un neurone artificiale, e ogni legame una connessione orientata dall'output di un neurone all'input di un altro. Ogni connessione può trasmettere informazione *(segnale)* da un neurone all'altro. Da questo modello originano architetture di calcolo e sistemi detti *connessionisti.* Neuroni e legami tipicamente hanno un fattore peso *(weight)* che si modifica *(aggiorna, update)* durante le fasi di apprendimento, e in tal modo rafforzano o indeboliscono l'intensità del segnale che li collega.
Col nome di *Apprendimento Profondo (Deep Learning)* si indica un modello ANN costituito da neuroni interconnessi e disposti in molteplici strati *(layer*). Modelli di Deep Learning sono stati applicati con successo a problemi di apprendimento per la Visione Artificiale e il Riconoscimento del Parlato. Nelle lezioni successive ci occuperemo in dettaglio dei modelli ANN e alcune loro applicazioni.

3.2 Alberi Decisionali (Decision Trees)

L'apprendimento è basato su un modello supervisionato usato anche in Statistica; Data Mining; Ricerca operativa, che fa parte dei *modelli previsionali (predictive models).* In questo formalismo viene usato un *albero di decisione,* appunto come modello previsionale per trarre conclusioni a partire da un insieme di osservazioni.

Si distinguono gli alberi decisionali in cui una variabile finale assume valori discreti *(alberi di classificazione)* da quelli in cui assume valori continui, tipicamente numeri reali *(alberi di regressione, regression trees).* Nel primo caso le foglie rappresentano *etichette di categorie (class labels),* e i rami *relazioni tra attributi (features)* che portano a quelle etichette. L'apprendimento tramite alberi di decisione è usato nel data mining per creare un modello che predice il valore di una variabile obiettivo sulla base di alcune variabili di input. Nell'*analisi decisionale* un albero può essere adottato per rappresentare in modo esplicito e visuale *le decisioni possibili,* e *il processo stesso che porta a una decisione (predizione).*



Un albero decisionale per la classificazione di dati meteo.

3.3 Analisi di Regressione (Regression Analysis)

Comprende un'ampia varietà di metodi statistici per stimare la relazione che intercorre tra le variabili di input e le *forme (features)* che ne descrivono l'andamento. Il metodo più comune è la *regressione lineare* in cui la forma che descrive i dati è un segmento di retta, i cui parametri sono stimati secondo il criterio di ottimizzazione *dei minimi quadrati.* Tale criterio si può estendere per tener conto di effetti numerici come *l'overfitting* e il *bias statistico.* Si intende per overfitting il processo che produce un'analisi che corrisponde troppo strettamente ai dati; il modello presumibilmente contiene più parametri di quelli che i dati osservati giustificano. Si intende per bias statistico una deviazione sistematica tra i dati osservati e i risultati dell'analisi.

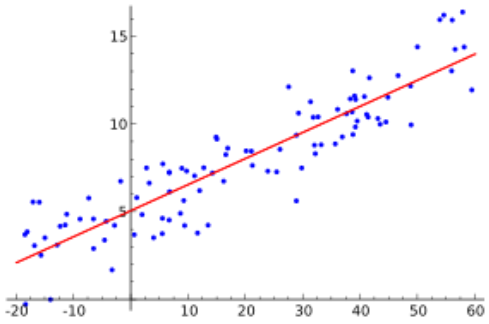
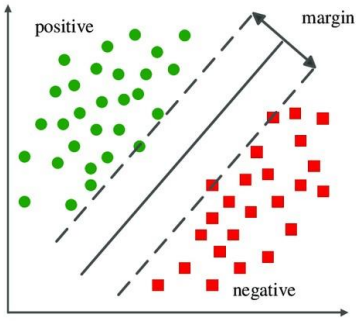


Illustrazione di una regressione lineare su un insieme di dati.

3.4 Support-Vector Machine (SVM)

Sono modelli ad apprendimento supervisionato applicati all'analisi dei dati per scopi di classificazione e regressione. Dato un insieme di dati di apprendimento (training set), ciascuno etichettato come appartenente a una tra due categorie, un algoritmo di apprendimento SVM costruisce un modello matematico che assegna nuovi esempi a una categoria o all'altra, ed è quindi un classificatore binario. Una SVM mette in corrispondenza gli esempi di apprendimento con punti nello spazio in modo da massimizzare l'ampiezza dell'intervallo *(margin)* tra le due categorie. Di conseguenza nuovi esempi sono mappati sullo stesso spazio, e se ne predice l'appartenenza a una categoria o all'altra a seconda di dove vanno a cadere, se da una parte o l'altra dell'intervallo.



Applicazione di un algoritmo SVM a un insieme di dati.

3.5 Algoritmi Genetici (Genetic Algorithms, GA)

Un modello genetico, e relativi algoritmi, è costituito da un processo di ricerca su grafo corredato da tecniche euristiche; è ispirato alla selezione naturale con l'obiettivo di generare soluzioni a problemi di ottimizzazione. Appartiene alla classe degli *algoritmi evolutivi (evolutionary).* Una *popolazione* è un insieme di *soluzioni candidate* per l'ottimizzazione *(organismi* o *fenotipi)*; tipicamente rappresentata da un insieme di stringhe binarie generate a caso, la popolazione viene fatta evolvere per ottenere soluzioni migliori. Ciascuna soluzione candidata contiene *proprietà* o *attributi (genotipi)*, che possono essere soggetti a *mutazioni (mutations)*. L'evoluzione consiste in un processo iterativo, in cui lo stato di una popolazione dopo ciascuna iterazione è chiamata *una generazione.* Di ciascuna generazione si valuta *l'adattamento (fitness)* di ogni individuo, che di solito è il valore di una *funzione obiettivo* rispetto al problema di ottimizzazione da risolvere. Gli individui con adattamento maggiore sono selezionati in ordine casuale, e ciascuno modificato per formare una nuova generazione per la successiva iterazione. L'algoritmo ha termine quando si raggiunge un livello di adattamento soddisfacente, oppure si raggiunge un numero prefissato di generazioni.

3.6 Reti Bayesiane (Bayesian Networks)

Una rete bayesiana è un *grafo aciclico orientato (Directed Acyclic Graph, DAG)* in cui:
- Le etichette dei nodi rappresentano *le variabili* in senso bayesiano: quantità osservabili; variabili latenti; parametri incogniti; ipotesi;
- Gli archi rappresentano *le relazioni di dipendenza condizionale (conditional dependencies)* tra le variabili e le distribuzioni di probabilità dei nodi-figlio rispetto ai valori dei nodi-padre.

Una rete bayesiana rappresenta la *distribuzione della probabilità congiunta* di un insieme di variabili, e in questo senso dà una rappresentazione completa della conoscenza, che comprende anche variabili non direttamente osservate. Nodi tra loro non connessi rappresentano variabili indipendenti l'una dall'altra. Ogni nodo è associato a una funzione di distribuzione di probabilità che prende in input un insieme di valori assunti dalle variabili del nodo genitore, a cui associa in output la probabilità condizionale della variabile stessa - opp. distribuzione di probabilità, se è il caso. Nel caso particolare in cui i nodi genitori rappresentano *variabili binarie (boolean)*, allora la probabilità si può rappresentare con una tabella avente un'entrata per ogni possibile combinazione dei valori associati al genitore.

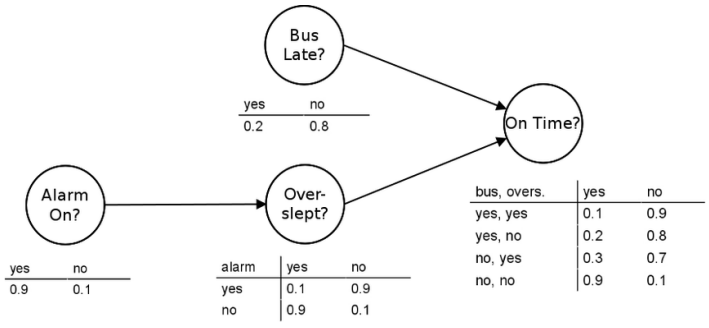


Illustrazione di una rete bayesiana a variabili binarie.

3.7 Reti bayesiane e apprendimento

Nell'ambito dell'IA le reti bayesiane sono usate per la *rappresentazione della conoscenza incerta* e del *ragionamento automatico (inferenza) in condizioni d'incertezza.*

Nell'apprendimento automatico le reti bayesiane hanno i seguenti scopi:
(a) Inferire valori di variabili non osservabili e loro relazioni;
(b) Apprendimento di parametri;
(c) Apprendimento strutturale.

Inferenza di valori di variabili non osservabili

Una rete si può usare per aggiornare la conoscenza su un sottoinsieme di variabili quando si osservano altre variabili dette *di evidenza.* Il processo di stima delle probabilità a-posteriori delle variabili non osservabili data l'evidenza è denominata *inferenza probabilistica.*

Apprendimento di parametri

Per la funzione di distribuzione condizionale di una variabile X a partire da quella dei suoi genitori nel DAG si è soliti adottare distribuzioni discrete di tipo gaussiano.

Apprendimento strutturale

Specificare la struttura di una rete è un problema complesso: può essere suggerita da un esperto del dominio di conoscenze; oppure, i parametri delle distribuzioni possono essere appresi direttamente da un insieme di dati. Dunque, può essere un problema di diretto interesse per il ML.