

INTELLIGENZA ARTIFICIALE

ANNO ACCADEMICO
2008-2009

Introduzione al Machine Learning
Apprendimento Induttivo: alberi di decisione

SOMMARIO

- Introduzione al Machine Learning
- Apprendimento Induttivo
- Spazio delle Ipotesi
- Alberi di Decisione

INTRODUZIONE AL MACHINE LEARNING

- ✿ Intuitivamente, un sistema è in grado di apprendere se, attraverso la sua attività, è in grado di migliorare le proprie prestazioni.
- ✿ Nell'IA, il miglioramento delle prestazioni coincide in generale con l'acquisizione di nuove conoscenze.

INTRODUZIONE AL MACHINE LEARNING

Qualsiasi cambiamento in un sistema che gli permetta di avere prestazioni migliori la seconda volta, nella ripetizione dello stesso compito o di un altro compito tratto dalla stessa popolazione.

(Simon, 1984)

INTRODUZIONE AL MACHINE LEARNING

A computer program is said to learn from experience **E** with respect to some class of tasks **T** and performance measure **P**, if its performance at tasks in **T**, as measured by **P**, **improves with experience E**.

(Mitchell, 1997)

DEFINIZIONI

- ✿ **Task T:** obiettivo del sistema
 - Giocare a dama
 - Guidare un autoveicolo
 - Riconoscere parole pronunciate

- ✿ **Experience E:** Insieme di addestramento dal quale apprendere
 - Partite giocate
 - Percorsi
 -

- ✿ **Performance measure P:** misura della capacità di eseguire il task
 - Numero di partite vinte
 - Numero di parole classificate correttamente

INTRODUZIONE AL MACHINE LEARNING

- ✿ Un elemento fondamentale dell'apprendimento è la capacità di valutare le proprie prestazioni, o almeno di accettare una valutazione dall'esterno.
- ✿ Senza una valutazione, infatti, non sarebbe possibile parlare di miglioramento.
- ✿ A sua volta, la valutazione delle prestazioni richiede la capacità di accettare un certo tipo di informazioni dall'ambiente esterno.

ESEMPI DI APPLICAZIONE

Speech Recognition

SPHINX system (Lee, 1989): riconoscimento fonemi

Guida autoveicolo

ALVINN system (Pomerlau, 1989)

Classificazione nuove strutture astronomiche

NASA: classificazione oggetti celesti (Fayyad et al., 1995)

Backgammon

TD-Gammon (Tesauro, 1992, 1995): apprendimento su 1 milione di partite giocate contro se stesso.

Checkers: apprendimento del gioco della dama

ESEMPI DI APPLICAZIONE

Gioco della Dama

Task T: giocare a dama

Performance measure P: % di partite vinte

Training Experience E: partite giocate contro se stesso

Riconoscimento di caratteri scritti a mano

Task T: riconoscere e classificare parole scritte a mano memorizzate come immagini

Performance measure P: % di parole correttamente classificate

Training Experience E: un database di parole scritte a mano insieme alla loro classificazione

Robot che guida un autoveicolo

Task T: guidare un autoveicolo su una strada pubblica

Performance measure P: distanza percorsa prima di un errore

Training Experience E: sequenza di immagini

METODI DI APPRENDIMENTO

✿ Apprendimento **supervisionato**

- Richiede che si apprenda una funzione partendo da esempi di input e output

✿ Apprendimento **non supervisionato**

- Richiede di imparare a riconoscere pattern o schemi nell'input senza alcuna indicazione specifica dei valori di uscita.

✿ Apprendimento per **rinforzo**

- L'agente apprende in base al rinforzo (ricompensa) ottenuto.

APPRENDIMENTO INDUTTIVO (INDUCTIVE LEARNING METHOD)

- ✿ Un algoritmo per l'apprendimento (deterministico) **supervisionato** riceve in ingresso il valore corretto in alcuni punti di una funzione sconosciuta.
- ✿ In base a tali valori deve cercare di ricostruire la funzione nel modo più fedele possibile.
- ✿ Definiamo **esempio** una coppia $(x, f(x))$, in cui x è l'input e $f(x)$ è l'output della funzione.

APPRENDIMENTO INDUTTIVO (INDUCTIVE LEARNING METHOD)

- ✻ Il compito dell'**inferenza induttiva pura** (o **induzione**) è il seguente:

Data una collezione di **esempi di f** , restituire una funzione **h** che approssima **f** .

- ✻ La funzione **h** prende il nome di **ipotesi**.

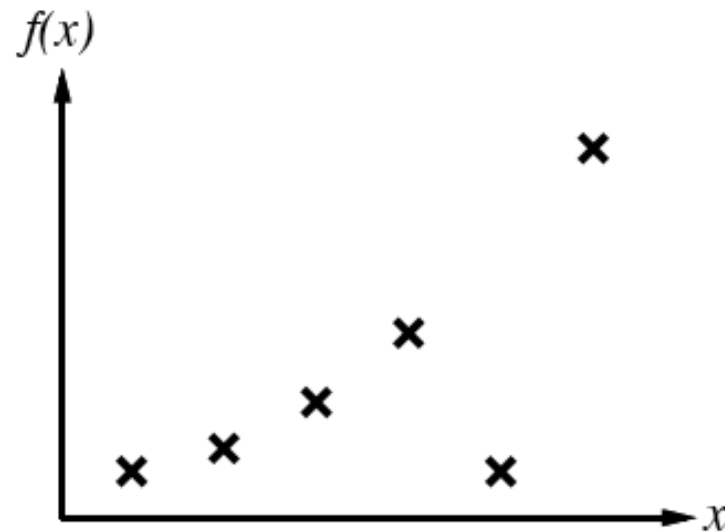
APPRENDIMENTO INDUTTIVO (INDUCTIVE LEARNING METHOD)

Il problema dell'induzione

- ✱ La difficoltà che si incontra in tale attività è dovuta al fatto che non è facile stabilire se una particolare h sia una buona approssimazione di f .
- ✱ Una buona ipotesi si potrà **generalizzare** bene, ossia potrà predire correttamente esempi che non ha ancora incontrato.

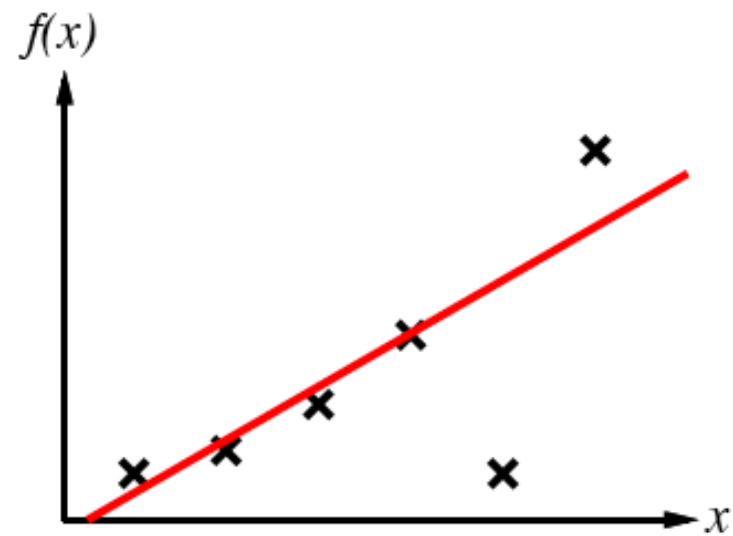
APPRENDIMENTO INDUTTIVO (INDUCTIVE LEARNING)

- ✿ Analizziamo la seguente figura che mostra alcune coppie di esempio $(x, f(x))$, in cui sia x che $f(x)$ sono numeri reali:

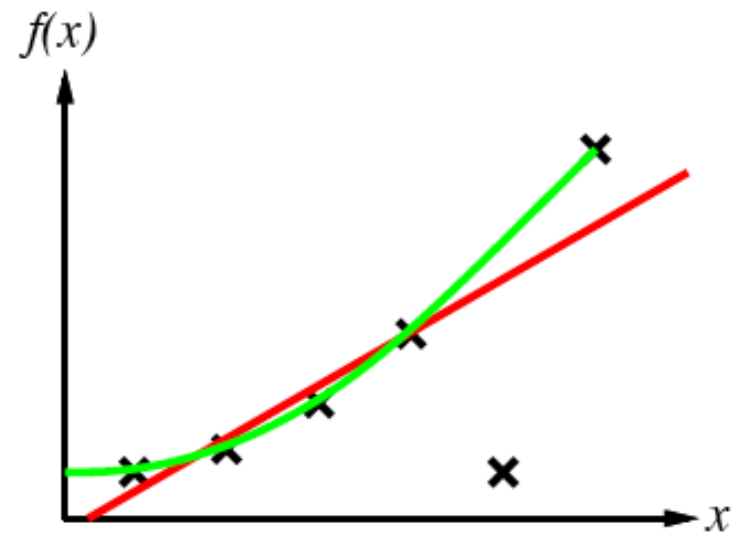


INDUCTIVE LEARNING

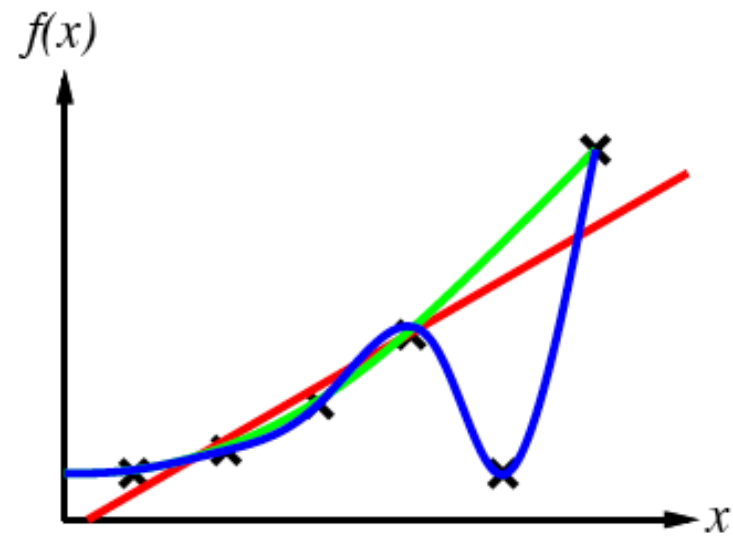
- ✿ Ci sono varie ipotesi possibili per tale insieme di dati.
Vediamone alcune:



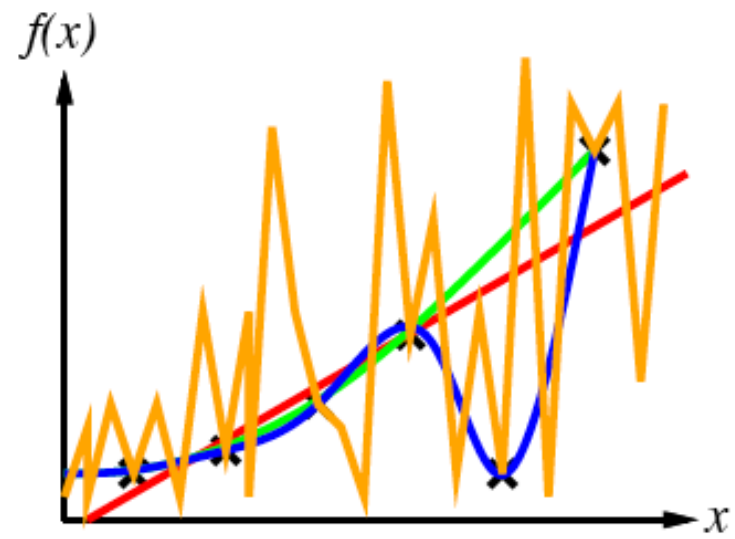
INDUCTIVE LEARNING



INDUCTIVE LEARNING

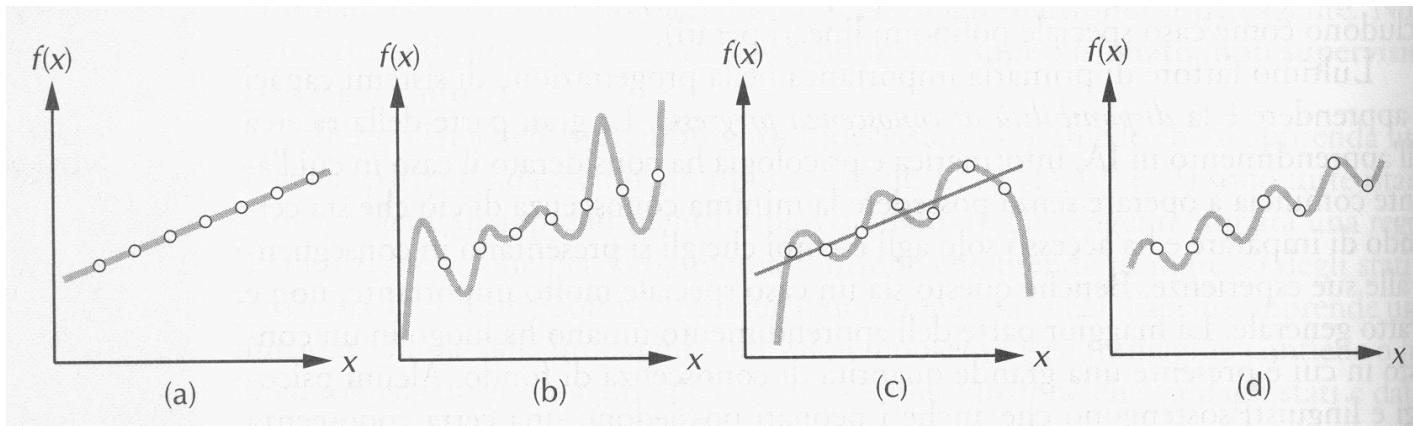


INDUCTIVE LEARNING



INDUCTIVE LEARNING

☼ Vediamo ancora un esempio, relativo alla figura che segue:



INDUCTIVE LEARNING

- ✱ Scegliamo come **spazio delle ipotesi H** (insieme di tutte le ipotesi considerate) l'insieme dei polinomi di grado massimo k .
- ✱ Nella figura (a) è mostrata un'ipotesi consistente lineare.
- ✱ La figura (b) riporta un polinomio di ordine superiore, anch'esso consistente con lo stesso insieme di dati.
- ✱ Ciò illustra il primo problema dell'apprendimento induttivo:

Come facciamo a scegliere tra più ipotesi,
tutte consistenti?

INDUCTIVE LEARNING

- ✿ Principio del **Rasoio di Ockham**:

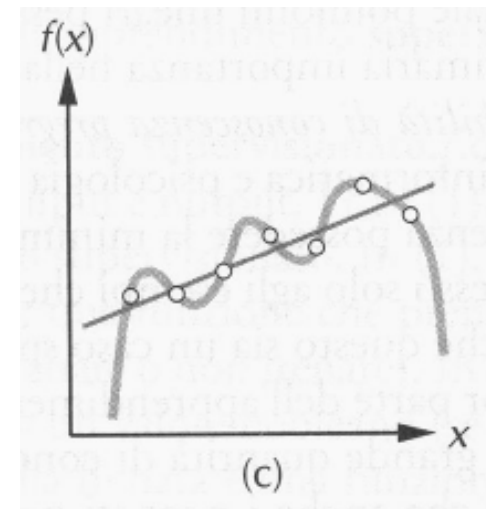
Si deve preferire l'ipotesi più semplice
consistente con i dati.

(entia non sunt multiplicanda praeter necessitatem)

- ✿ Non è facile definire la “semplicità” di una ipotesi, ma è certamente ragionevole affermare che, considerando ad esempio il caso precedente, un polinomio di grado 1 sia più semplice di un polinomio di grado 12.

INDUCTIVE LEARNING

- ✱ Nella figura (c) è mostrato un secondo insieme di dati.
- ✱ In questo caso non esiste un'ipotesi consistente lineare.
- ✱ E' infatti richiesto un polinomio di ordine superiore (di grado 6) per arrivare ad una consistenza esatta.
- ✱ Una soluzione migliore potrebbe essere una semplice linea retta che, pur non essendo esattamente consistente, potrebbe comunque fornire previsioni ragionevoli.



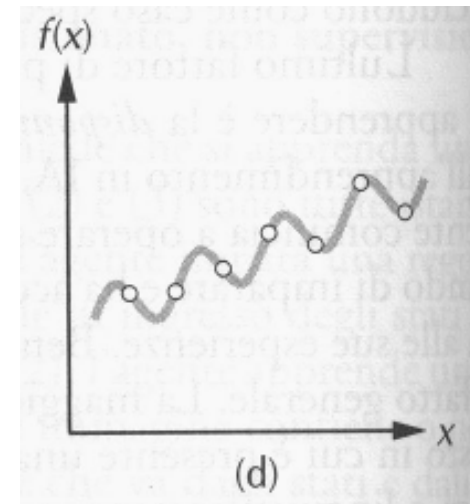
INDUCTIVE LEARNING

- ✿ Scegliere l'ipotesi lineare nella figura (c) significa accettare la possibilità che la vera funzione non sia deterministica o, ciò che è sostanzialmente equivalente, che i veri input non siano completamente osservati.
- ✿ Dobbiamo dunque ricordare che:

Nel caso di funzioni non deterministiche occorre accettare un inevitabile compromesso tra la complessità dell'ipotesi e il grado di corrispondenza con i dati.

INDUCTIVE LEARNING

- ☼ Dobbiamo poi tener presente che la possibilità o l'impossibilità di trovare un'ipotesi semplice e consistente dipende fortemente dallo spazio delle ipotesi prescelto.
- ☼ La figura (d) mostra che gli stessi dati della figura (c) possono corrispondere esattamente ad una semplice funzione della forma: $ax + b + c \sin x$.
- ☼ Questo esempio dimostra l'importanza della scelta dello spazio delle ipotesi.



ALBERI DI DECISIONE

- ✿ L'induzione di alberi di decisione è una delle forme di apprendimento più semplici e ciononostante di maggior successo.
- ✿ Costituisce dunque una buona introduzione al campo dell'apprendimento induttivo ed è semplice da implementare.

ALBERI DI DECISIONE

- ✿ Un albero di decisione prende come ingresso un oggetto o una situazione descritta da un insieme di **attributi** e restituisce una “decisione”.
- ✿ Ci concentreremo sul caso della **classificazione booleana** (**attributi discreti**, esempio classificato come **positivo** o **negativo**).
- ✿ Ogni nodo interno dell'albero è connesso ad un test sul valore di un attributo.
- ✿ Ogni foglia dell'albero specifica una possibile decisione, ovvero il valore predetto di uscita per un certo input.

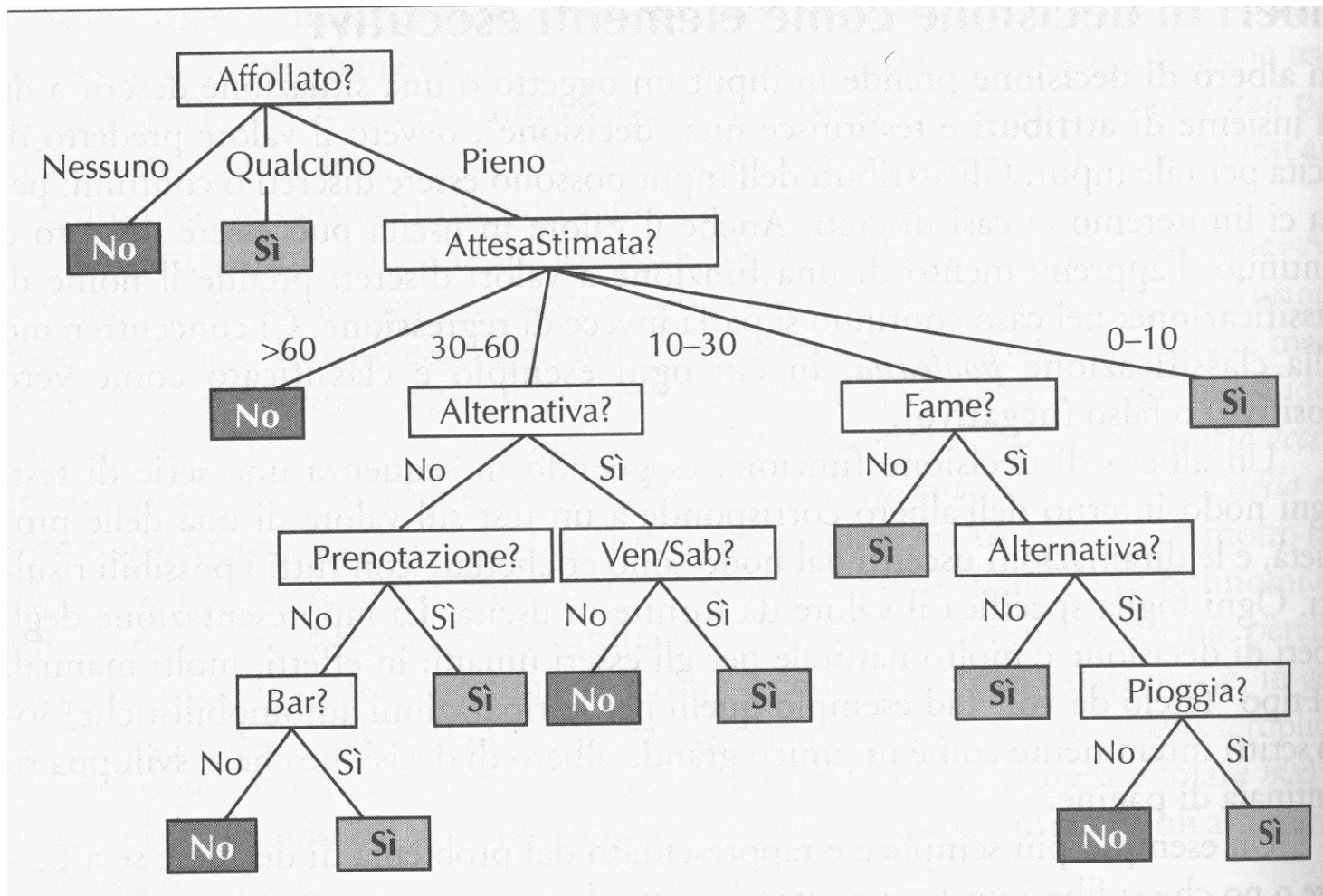
ESEMPIO PER IL DOMINIO DEL RISTORANTE

- ✿ Il problema è quello di decidere se aspettare o no che si liberi un tavolo quando si va al ristorante.
- ✿ In questo caso lo scopo è apprendere una definizione del **predicato obiettivo** “Aspettiamo”, dove la definizione è espressa sotto forma di albero di decisione.
- ✿ Nel formulare questo problema come un problema di apprendimento dobbiamo anzitutto decidere quali proprietà, o attributi, sono disponibili per descrivere esempi nel dominio.

ESEMPIO PER IL DOMINIO DEL RISTORANTE

- ✿ Supponiamo di avere deciso per la lista di attributi seguente:
 - **Alternativa**: se c'è un ristorante nei paraggi
 - **Bar**: se il ristorante ha un'area bar per l'attesa
 - **Ven/Sab**: vero di venerdì e i sabato
 - **Fame**: se siamo affamati
 - **Affollato**: quante persone sono presenti nel ristorante
 - **Prezzo**: categoria di costo del ristorante
 - **Pioggia**: se fuori sta piovendo
 - **Prenotazione**: se abbiamo prenotato
 - **Tipo**: tipo di ristorante (italiano, francese, fast-food, ecc.)
 - **Attesa stimata**: stima del tempo di attesa

ALBERO DI DECISIONE (DI RUSSELL!) PER IL DOMINIO DEL RISTORANTE



ALBERO DI DECISIONE (DI RUSSELL!) PER IL DOMINIO DEL RISTORANTE

- ✿ Si osservi che l'albero non si serve degli attributi **Prezzo** e **Tipo**, in quanto si possono considerare irrilevanti considerati i dati osservati.
- ✿ In termini logici l'albero di decisione può essere espresso come congiunzione di implicazioni corrispondenti ciascuna a uno dei cammini terminanti in un nodo **Sì**.
- ✿ Ad esempio, il cammino per il caso di un ristorante pieno di clienti, con un tempo di attesa stimato di 10-30 minuti nel caso in cui l'agente non sia affamato può essere espresso mediante la formula logica:

$$\forall r \text{ Affollato}(r, \text{Pieno}) \wedge \text{AttesaStimata}(r, 10-30) \wedge \text{Fame}(r, \text{No}) \Rightarrow \text{Aspettiamo}(r)$$

INDUZIONE DI ALBERI DI DECISIONE A PARTIRE DA ESEMPI

- ✿ Un esempio per un **albero di decisione booleano** è descritto dai valori degli **attributi** e dal valore del **predicato obiettivo**.
- ✿ Chiamiamo il valore del predicato obiettivo **classificazione** dell'esempio.
- ✿ Se il predicato è vero per un certo esempio, chiamiamo tale esempio **positivo**, mentre in caso contrario diciamo che è **negativo**.
- ✿ La tabella che segue mostra un insieme di esempi X_1, \dots, X_{12} per il dominio del ristorante.
- ✿ L'insieme completo degli esempi è chiamato insieme di training.

ESEMPI PER IL DOMINIO DEL RISTORANTE

Esempio	Attributi										Obiettivo
	<i>Alt</i>	<i>Bar</i>	<i>Ven</i>	<i>Fame</i>	<i>Affoll</i>	<i>Prezzo</i>	<i>Piogg</i>	<i>Prenot</i>	<i>Tipo</i>	<i>Attesa</i>	<i>Aspettiamo</i>
X_1	<i>Sì</i>	<i>No</i>	<i>No</i>	<i>Sì</i>	<i>Qualcuno</i>	<i>\$\$\$</i>	<i>No</i>	<i>Sì</i>	<i>Francese</i>	<i>0-10</i>	<i>Sì</i>
X_2	<i>Sì</i>	<i>No</i>	<i>No</i>	<i>Sì</i>	<i>Pieno</i>	<i>\$</i>	<i>No</i>	<i>No</i>	<i>Thai</i>	<i>30-60</i>	<i>No</i>
X_3	<i>No</i>	<i>Sì</i>	<i>No</i>	<i>No</i>	<i>Qualcuno</i>	<i>\$</i>	<i>No</i>	<i>No</i>	<i>Fast-food</i>	<i>0-10</i>	<i>Sì</i>
X_4	<i>Sì</i>	<i>No</i>	<i>Sì</i>	<i>Sì</i>	<i>Pieno</i>	<i>\$</i>	<i>Sì</i>	<i>No</i>	<i>Thai</i>	<i>10-30</i>	<i>Sì</i>
X_5	<i>Sì</i>	<i>No</i>	<i>Sì</i>	<i>No</i>	<i>Pieno</i>	<i>\$\$\$</i>	<i>No</i>	<i>Sì</i>	<i>Francese</i>	<i>>60</i>	<i>No</i>
X_6	<i>No</i>	<i>Sì</i>	<i>No</i>	<i>Sì</i>	<i>Qualcuno</i>	<i>\$\$</i>	<i>Sì</i>	<i>Sì</i>	<i>Italiano</i>	<i>0-10</i>	<i>Sì</i>
X_7	<i>No</i>	<i>Sì</i>	<i>No</i>	<i>No</i>	<i>Nessuno</i>	<i>\$</i>	<i>Sì</i>	<i>No</i>	<i>Fast-food</i>	<i>0-10</i>	<i>No</i>
X_8	<i>No</i>	<i>No</i>	<i>No</i>	<i>Sì</i>	<i>Qualcuno</i>	<i>\$\$</i>	<i>Sì</i>	<i>Sì</i>	<i>Thai</i>	<i>0-10</i>	<i>Sì</i>
X_9	<i>No</i>	<i>Sì</i>	<i>Sì</i>	<i>No</i>	<i>Pieno</i>	<i>\$</i>	<i>Sì</i>	<i>No</i>	<i>Fast-food</i>	<i>>60</i>	<i>No</i>
X_{10}	<i>Sì</i>	<i>Sì</i>	<i>Sì</i>	<i>Sì</i>	<i>Pieno</i>	<i>\$\$\$</i>	<i>No</i>	<i>Sì</i>	<i>Italiano</i>	<i>10-30</i>	<i>No</i>
X_{11}	<i>No</i>	<i>No</i>	<i>No</i>	<i>No</i>	<i>Nessuno</i>	<i>\$</i>	<i>No</i>	<i>No</i>	<i>Thai</i>	<i>0-10</i>	<i>No</i>
X_{12}	<i>Sì</i>	<i>Sì</i>	<i>Sì</i>	<i>Sì</i>	<i>Pieno</i>	<i>\$</i>	<i>No</i>	<i>No</i>	<i>Fast-food</i>	<i>30-60</i>	<i>Sì</i>

INDUZIONE DI ALBERI DI DECISIONE A PARTIRE DA ESEMPI

- ✿ Il problema da risolvere è trovare un albero di decisione che sia consistente con l'insieme di training.
- ✿ Una soluzione semplice potrebbe essere la seguente:
 - costruiamo un albero di decisione con un cammino completo fino a una foglia per ogni esempio, in cui in tale cammino si verifica che ciascun attributo possieda lo stesso valore assunto nell'esempio e la foglia ha la stessa classificazione dell'esempio.
- ✿ E' ovvio che, qualora lo stesso esempio venga presentato nuovamente all'albero, esso sarà in grado di fornire la classificazione corretta. Purtroppo però negli altri casi non avrà molto da dire!

INDUZIONE DI ALBERI DI DECISIONE A PARTIRE DA ESEMPI

- ✿ Il problema di questo albero banale è che si limita a memorizzare le informazioni.
- ✿ Non estrae alcuno schema dagli esempi e quindi non ci si può attendere che sia in grado di estrapolare al caso di esempi che non ha mai visto.
- ✿ Estrarre uno schema significa essere in grado di descrivere un gran numero di casi in maniera concisa.

INDUZIONE DI ALBERI DI DECISIONE A PARTIRE DA ESEMPI

- ✿ Non solo dobbiamo cercare un albero di decisione che si accordi con gli esempi, ma dobbiamo anche cercarne uno che sia conciso (in sostanza dobbiamo applicare il principio del rasoio di Occam).
- ✿ Sfortunatamente il problema di trovare il più piccolo albero di decisione è intrattabile, ma con delle semplici euristiche possiamo fare un buon lavoro nel trovarne uno abbastanza piccolo.

ALGORITMO DI APPRENDIMENTO

1. Se ci sono esempi sia positivi sia negativi, scegliere l'attributo che li suddivide meglio.
2. Se i rimanenti esempi sono tutti positivi o tutti negativi, allora abbiamo finito.
3. Se non rimane alcun esempio, vuol dire che non è stato osservato nessun esempio con quei valori per gli attributi. Possiamo restituire un valore di default.
4. Se non rimane nessun attributo, ma ci sono ancora esempi sia positivi che negativi, abbiamo un problema (c'è **rumore** nei dati).

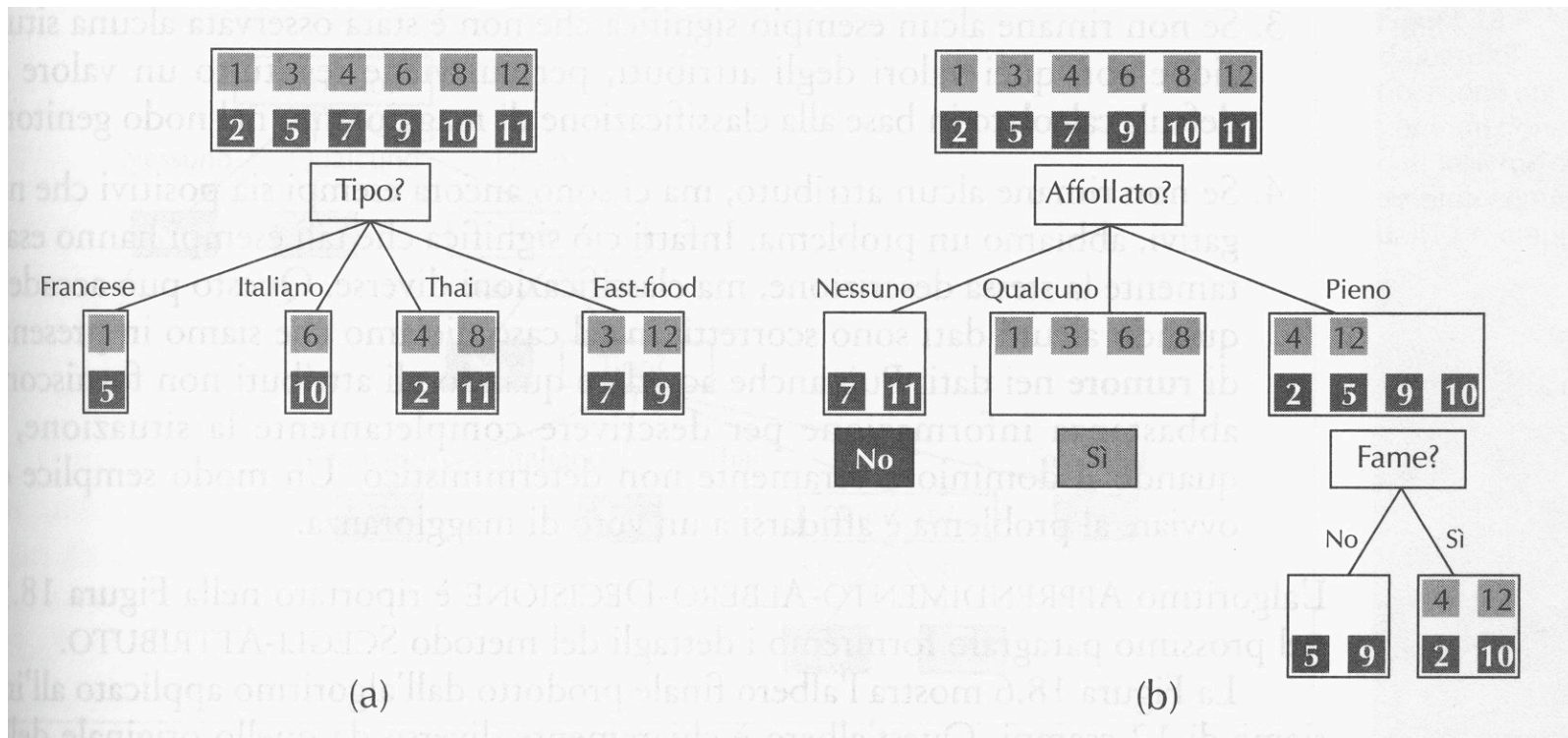
ALGORITMO DI APPRENDIMENTO

```
function APPRENDIMENTO-ALBERO-DECISIONE(esempi, attrib, default)  
    returns un albero di decisione  
  
inputs: esempi, un insieme di esempi  
        attrib, un insieme di attributi  
        default, valore di default del predicato obiettivo  
  
if esempi è vuoto then return default  
else if tutti gli esempi hanno la stessa classificazione then return la classificazione  
else if attrib è vuoto then return VALORE-MAGGIORANZA(esempi)  
else  
    best ← SCEGLI-ATTRIBUTO(attrib, esempi)  
    albero ← un nuovo albero di decisione con test alla radice best  
    m ← VALORE-MAGGIORANZA(esempi)  
    for each valore vi di best do  
        esempii ← {elementi di esempi con best = vi}  
        sottoalb ← APPRENDIMENTO-ALBERO-DECISIONE(esempii, attrib – best, m)  
        aggiungi un ramo all'albero con etichetta vi e sottoalbero sottoalb  
    return albero
```

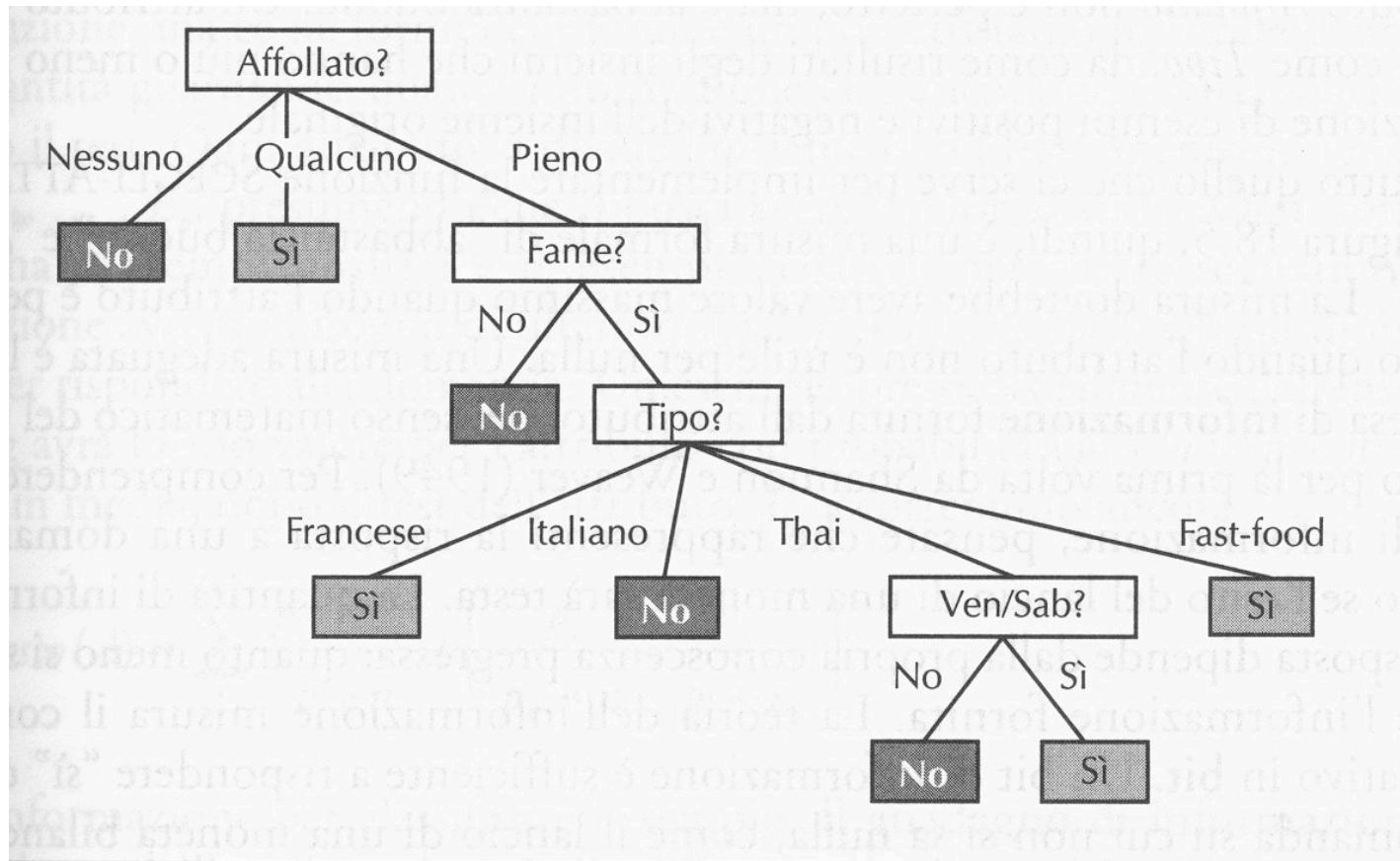
SCelta DI UN ATTRIBUTO

La divisione su “Tipo” non è di molto aiuto.

La divisione su “Affollato” è invece efficace per separare gli esempi positivi dai negativi.



ALBERO DI DECISIONE APPRESO PER INDUZIONE



VALUTAZIONE DELLE PRESTAZIONI

- ✿ Un algoritmo di apprendimento è buono se produce ipotesi che fanno un buon lavoro nel prevedere la classificazione di esempi non visti.
- ✿ Ovviamente una predizione è buona se si dimostra vera, quindi possiamo valutare la qualità di un'ipotesi confrontando la sua predizione con la classificazione corretta una volta che essa sia nota.
- ✿ Facciamo questo su un insieme di esempi noto come **insieme di test**.

VALUTAZIONE DELLE PRESTAZIONI

- ✿ Qualora effettuassimo l'addestramento su tutti gli esempi a nostra disposizione non avremmo poi i dati necessari per effettuare il test.

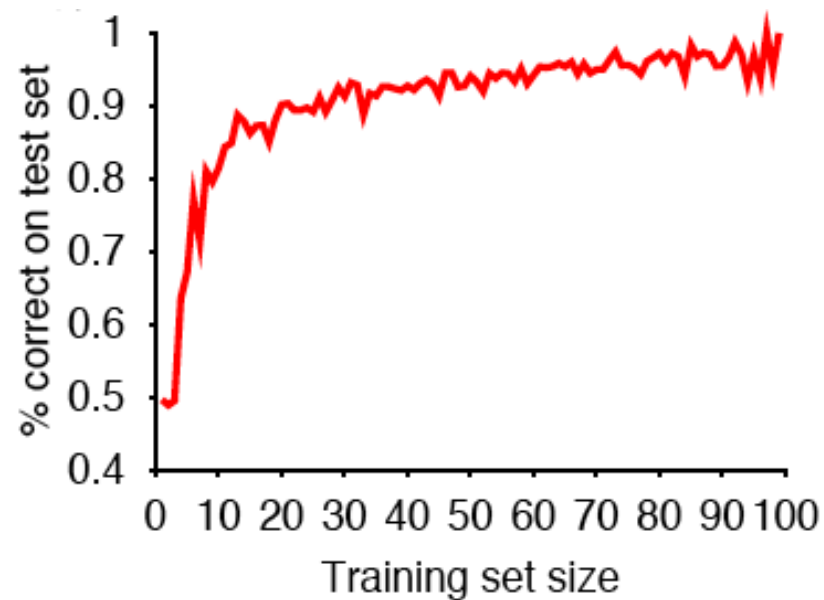
- ✿ E' perciò più conveniente adottare la seguente metodologia:
 1. Si raccoglie un grande insieme di esempi.
 2. Lo si divide in due: **insieme di addestramento** e **insieme di test**.
 3. Si usa l'algoritmo di apprendimento sul primo insieme.
 4. Si misura la percentuale di esempi nell'insieme di test che viene classificata correttamente dall'ipotesi generata nel passo 3.
 5. Si ripetono i passi da 1 a 4 con insiemi di addestramento di dimensioni diverse e, per ciascuna dimensione, con diversi insiemi selezionati casualmente.

VALUTAZIONE DELLE PRESTAZIONI

- ✿ Il risultato di queste operazioni è un insieme di dati che può essere elaborato per fornire la qualità media della predizione come una funzione della dimensione dell'insieme di addestramento.
- ✿ Tale funzione può essere tracciata su un grafico ottenendo la cosiddetta **curva di apprendimento** per l'algoritmo nel particolare dominio.

VALUTAZIONE DELLE PRESTAZIONI

- ✿ La figura seguente è relativa alla curva di apprendimento dell'algoritmo per gli alberi di decisione su 100 esempi generati casualmente nel dominio del ristorante. Il grafo riassume 20 tentativi:



VALUTAZIONE DELLE PRESTAZIONI

- ✿ Si osservi che al crescere dell'insieme di addestramento la qualità della predizione migliora (tali curve sono chiamate anche **grafi felici**).
- ✿ Questo comportamento rappresenta un buon segno: vuol dire che c'è una qualche regolarità nei dati e che l'algoritmo di apprendimento la sta estraendo.

VALUTAZIONE DELLE PRESTAZIONI

- ✿ L'idea fondamentale della metodologia consiste nel tenere **separati** i dati di addestramento da quelli di test, per la stessa ragione per cui i risultati di un esame non sarebbero una buona misura della qualità degli studenti se questi ultimi vedessero in anticipo il testo dell'esame.
- ✿ La metodologia che consiste nel suddividere a caso gli esempi in un insieme di addestramento e in un insieme di test è corretta quando **ciascuna esecuzione è indipendente dalle altre**: in tal caso nessuna esecuzione può "barare" e dire alle altre esecuzioni quali sono le risposte esatte.

VALUTAZIONE DELLE PRESTAZIONI

Attenzione: è l'autore dell'algoritmo che può barare!

- ✿ Purtroppo è molto facile cadere nella tentazione di dare una “sbirciata” ai dati di test (**peeking**).
- ✿ Il peeking si verifica in questo modo:
 - si generano diverse ipotesi con varie versioni dell'algoritmo (variandone ad es. dei parametri)
 - si misurano le loro prestazioni sull'insieme di test
 - si sceglie l'ipotesi che fornisce le migliori predizioni

VALUTAZIONE DELLE PRESTAZIONI

Si è verificato un peeking!

- ✿ E' accaduto che l'ipotesi è stata selezionata **in base alle sue prestazioni sull'insieme di test.**
- ✿ L'informazione che avrebbe dovuto rimanere confinata in tale insieme si è, per così dire, "infiltrata" nell'algoritmo di apprendimento.

VALUTAZIONE DELLE PRESTAZIONI

Morale:

- ✿ Ogni processo che prevede il confronto di varie ipotesi su un insieme di test dovrebbe usare un **nuovo** insieme di test per misurare l'ipotesi selezionata.
- ✿ Nella pratica questo è troppo difficile, per cui si continua ad effettuare esperimenti su insiemi di esempi già “sporcati”.

ESEMPIO DI APPLICAZIONE

- ✿ Nel 1986 la BP mise in funzione un sistema esperto chiamato **GASOIL** per progettare sistemi di separazione di petrolio e gas per le piattaforme petrolifere d'alto mare.
- ✿ La separazione del petrolio dal gas viene effettuata presso la bocca del pozzo da un sistema grande, complicato e costoso la cui struttura dipende da numerosi attributi tra cui la proporzione tra gas, petrolio e acqua, l'intensità del flusso, la pressione, la densità, la viscosità, la temperatura e la tendenza alla produzione di cere.

ESEMPIO DI APPLICAZIONE

- ✿ A quel tempo **GASOIL** era il più grande sistema esperto commerciale del mondo, con le sue circa 2500 regole.
- ✿ Costruire un tale sistema a mano avrebbe richiesto circa 10 anni/uomo.
- ✿ Usando metodi di apprendimento di alberi di decisione applicati ad un database di progetti esistenti, il sistema è stato sviluppato in 100 giorni/uomo (Michie, D. "Current developments in expert systems", in *Proc. of the 2nd Australian Conference on Applications of Expert Systems*, Sydney, Australia, 1986, pp. 163-182).
- ✿ Molti sostengono che **GASOIL** superi gli esperti umani e che abbia fatto risparmiare alla BP milioni di dollari.