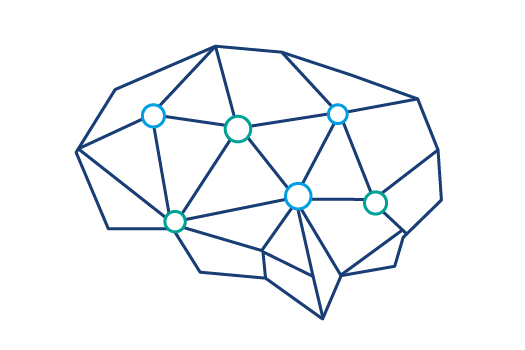
Corso di Intelligenza Artificiale. Prof. Filippo Neri

Anno Accademico 2020-2021

Work Projects

- Modeling with decision trees -

****

## 

Antonio Romano

Giuseppe Riccio

Salvatore Pernice  
Salvatore Ruggiero

**INDICE**

[**PREFAZIONE** 3](#_Toc70609043)

[**ESERCITAZIONE 1: MODELING WITH DECISION TREES** 4](#_Toc70609044)

[**ESERCIZIO 1** 4](#_Toc70609045)

[**1.1 Introduzione alla costruzione di un albero di decisione.** 4](#_Toc70609046)

[**1.2 Dataset Echocardiogram** 5](#_Toc70609047)

[**1.3 Dataset Mushrooms** 7](#_Toc70609048)

[**ESERCIZIO 2** 9](#_Toc70609049)

[**2.1 Perché' un agente in grado di apprendere Alberi di Decisione è considerato Intelligente?** 9](#_Toc70609050)

[**2.2 Cosa si intende con il termine apprendimento quando si utilizza il DT learning?** 10](#_Toc70609051)

[**ESERCIZIO 3** 11](#_Toc70609052)

[**3.1 Spiegare l’algoritmo A\*** 11](#_Toc70609053)

[**3.2 Descrivere i casi in cui si applica e cosa si intende per euristica ammissibile in A\*** 11](#_Toc70609054)

[**3.3 Mostrare un esempio di applicazione di A\* su un problema di navigazione stradale** 12](#_Toc70609055)

[**ESERCIZIO 4** 15](#_Toc70609056)

[**4.1 Descrivere con un esempio come si utilizza un albero di decisione** 15](#_Toc70609057)

[**ESERCIZIO 5** 17](#_Toc70609058)

[**5.1 Descrivere con un esempio il processo di apprendimento di un albero di decisione** 17](#_Toc70609059)

[**ESERCIZIO 6** 19](#_Toc70609060)

[**6.1 Cosa sono: lo spazio degli stati, il grafo degli stati e l’albero di ricerca (Tree Search)** 19](#_Toc70609061)

[**6.2 Descrivere in modo sintetico il meta algoritmo generale di costruzione di un Tree Search** 19](#_Toc70609062)

[**6.3 Spiegare in modo sintetico le caratteristiche degli algoritmi Tree Search visti a lezione** 19](#_Toc70609063)

**PREFAZIONE**

*La stesura della seguente trattazione è mirata alla documentazione di realizzazione di esercitazioni assegnate. Le tracce delle esercitazioni:*

**Esercitazione 1**

***Esercizio 1***

1) Eseguire sul proprio elaboratore (non riscrivere) il codice riportato nel Capitolo 7 'Modeling with Decision Trees' [wb1], pp 142 - 165, riflettere sui risultati ottenuti dall'esecuzione del codice e su quanto discusso in classe. Avete quindi ora un ambiente software disponibile per fare qualche esperienza pratica su DT learning

2) Scaricare un dataset da UC Irvine ML Repository a cui applicare DT Learning (codice del punto 1) e commentare per iscritto i risultati ottenuti. Un possibile dataset da utilizzare è chiamato Mushrooms.

3) Cambiare la percentuale di dati nell'insieme di training e di test (10%-90%, 20%-80%, …) e creare un grafico con le performance di apprendimento

Scrivere un report di max 6 pagine sulla vostra esperienza per i punti su indicati

***Esercizio 2***

Perché' un agente in grado di apprendere Alberi di Decisione è considerato Intelligente? Cosa si intende con il termine apprendimento quando si utilizza il DT learning?

***Esercizio 3***

Spiegare l’algoritmo A\*, descrivere i casi in cui si applica, cosa si intende per euristica ammissibile in A\*, mostrare un esempio di applicazione di A\* su un problema di navigazione stradale.

***Esercizio 4***

Descrivere con un esempio come si utilizza un albero di decisione.

***Esercizio 5***

Descrivere con un esempio il processo di apprendimento di un albero di decisione.

***Esercizio 6***

Cosa sono: lo spazio degli stati, il grafo degli stati e l’albero di ricerca (Tree Search). Descrivere in modo sintetico il meta algoritmo generale di costruzione di un Tree Search. Spiegare in modo sintetico le caratteristiche degli algoritmi Tree Search visti a lezione.

**ESERCITAZIONE 1:   
MODELING WITH DECISION TREES**

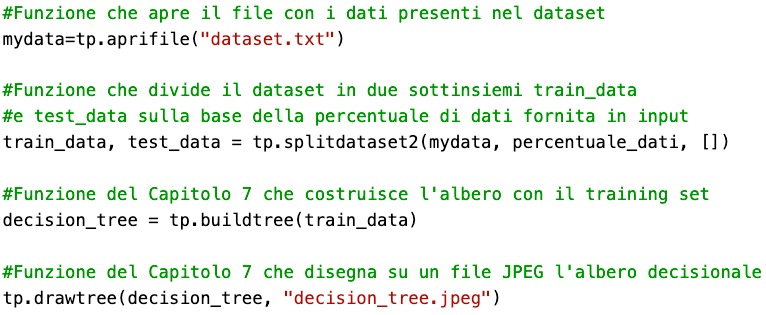
**ESERCIZIO 1**

**1.1 Introduzione alla costruzione di un albero di decisione.**

Gli alberi di decisione (o alberi decisionali) sono una rappresentazione dell'apprendimento nel **machine learning**. Un albero di decisione è un sistema con ***n*** variabili in input e***m*** variabili in output. Le variabili in input (**attributi**) sono derivate dall'osservazione dell'ambiente. Le variabili in output, invece, identificano la decisione/azione da intraprendere. (Cfr. Es.1.4 Successivamente).

Grazie al supporto del codice riportato nel Capitolo 7 'Modeling with Decision Trees' si è riflettuto sui risultati ottenuti dall’esecuzione del codice con l’utilizzo di due dataset prelevati dall’UC Irvine ML Repository. Partendo dal dataset iniziale costruiamo l’albero di decisione su una percentuale di dati desiderata (ad esempio il 60%) questo insieme di dati prende il nome di **training set**, mentre i dati rimanenti saranno utilizzati per la fase di test dell’albero decisionale ottenuto e per questo l’insieme viene chiamato **test set**.

Con il codice fornito nel Capitolo 7 e quello implementato dal nostro gruppo, dunque, i passi logici da seguire nell’esecuzione sarà il seguente:



**Figura 1**: Costruzione dell'albero di decisione

Quindi, dopo la costruzione dell’albero si passa alla fase di test per valutarne la sua performance inserendo i dati del test set all’interno della funzione (***learningcurve()*** sviluppata da noi) e utilizzando percentuali via via crescenti (in maniera del tutto automatica all’interno della funzione stessa, il testing set viene incrementato dal 10% fino al 90%) dell’insieme di dati di test, la funzione restituisce un grafico che ci permetterà di capire per ogni percentuale di dati test utilizzati quanti casi vengono trattati correttamente nell’albero decisionale e quindi, capire se l’albero costruito generalizza correttamente il problema. Proseguendo i passi di esecuzione del codice della Figura 1, avremo che per costruire il grafico di cui abbiamo parlato precedentemente occorrerà aggiungere la seguente riga di codice:

**Figura 2**: Funzione che disegna il grafico delle performance

L’informazione che fornisce il grafico è particolarmente indicativa e viene spesso espressa in questo modo:

***Accuratezza = % casi correttamente trattati***

Da cui può essere ricavato anche il seguente parametro di bontà dell’albero:

***ErrorRate = 1 – Accuratezza***

La caratteristica *Training Set Size - % Correct On Test Set* a seconda del suo andamento assume diversi significati:

*Realizable  
Redundant  
No-Realizable*

*% Correct*

*# Examples*

1

**Figura 3:** Tipici andamenti della Learning Curve

- Se la caratteristica si avvicina a quella ***realizzabile***allora si è appreso il modello perfetto.

- Se la caratteristica si avvicina a quella ***non realizzabile***non migliora più anche aumentando i dati. In questo caso o il problema non è deterministico oppure abbiamo rappresentato il problema con una tecnica non adeguata.

- Se la caratteristica si avvicina a quella ***ridondante***vuol dire che abbiamo creato uno spazio delle possibili ipotesi così grande che la tecnica ci mette tempo per trovare l’albero adatto.

Per il nostro caso di studio abbiamo selezionato dall’UCI i seguenti dataset:

- **Mushrooms** <https://archive.ics.uci.edu/ml/datasets/Mushroom>

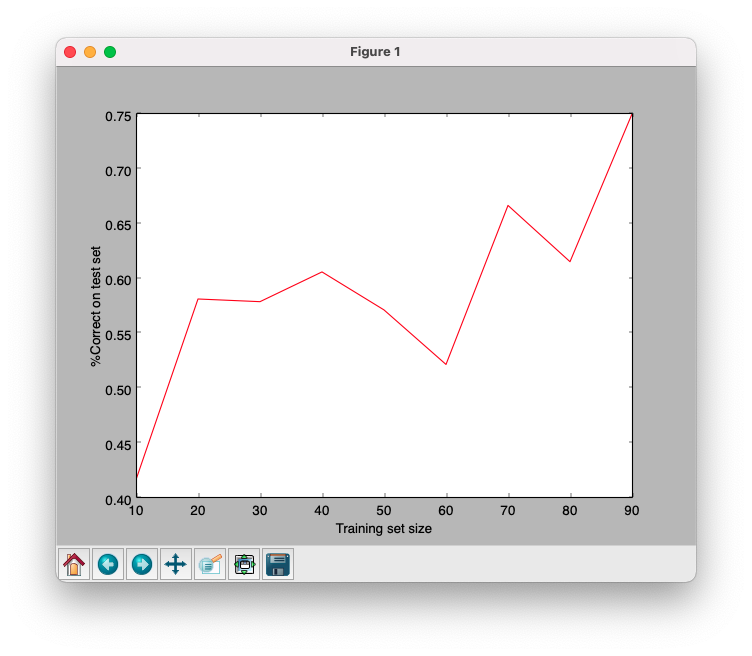
- **Echocardiogram** <https://archive.ics.uci.edu/ml/datasets/Echocardiogram>

**NB:** I dataset sono stati scelti sulla base della diversificazione nella loro “correttezza”.

**1.2 Dataset Echocardiogram**

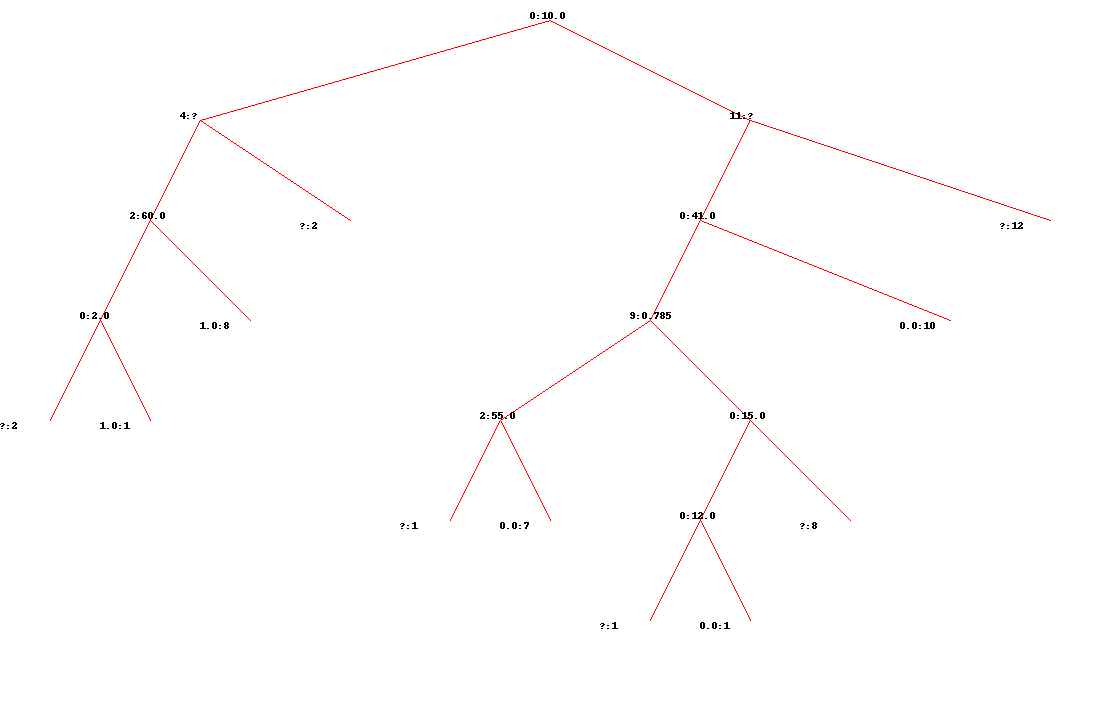
**Scopo:** Nel dataset sono presenti i dati di tutti i pazienti che hanno sofferto di attacchi di cuore in passato. Alcuni sono ancora vivi e altri no. Le variabili di sopravvivenza e ancora in vita, se prese insieme, indicano se un paziente è sopravvissuto per almeno un anno dopo l'attacco di cuore. Il problema da affrontare è prevedere correttamente se il paziente sopravvivrà o meno.

Il dataset è caratterizzato da ben 12 attributi e 132 istanze.

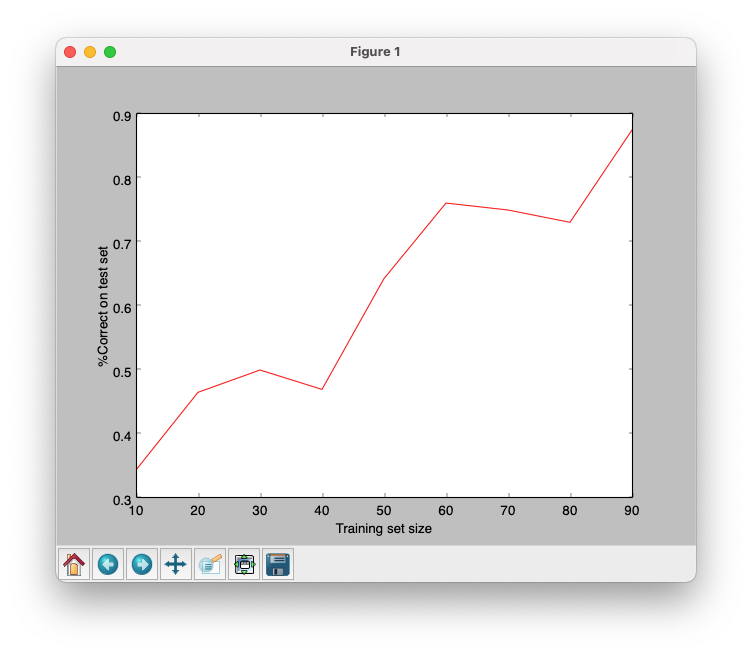
***Echocardiogram training\_set=40%, test\_set=60%***

Usando il 40% dati per il training (ovvero 53 istanze), abbiamo che l’albero (**Figura 5**) ha profondità massima 6, ed un grafico delle performance tendenzialmente crescente fino ad un massimo di circa 0.75 con alcuni punti in corrispondenza del 40%-60%, in cui la performance si riduce. (**Figura 4**)

**Figura 4:** Performance training\_set = 40% e test\_set = 60%

****

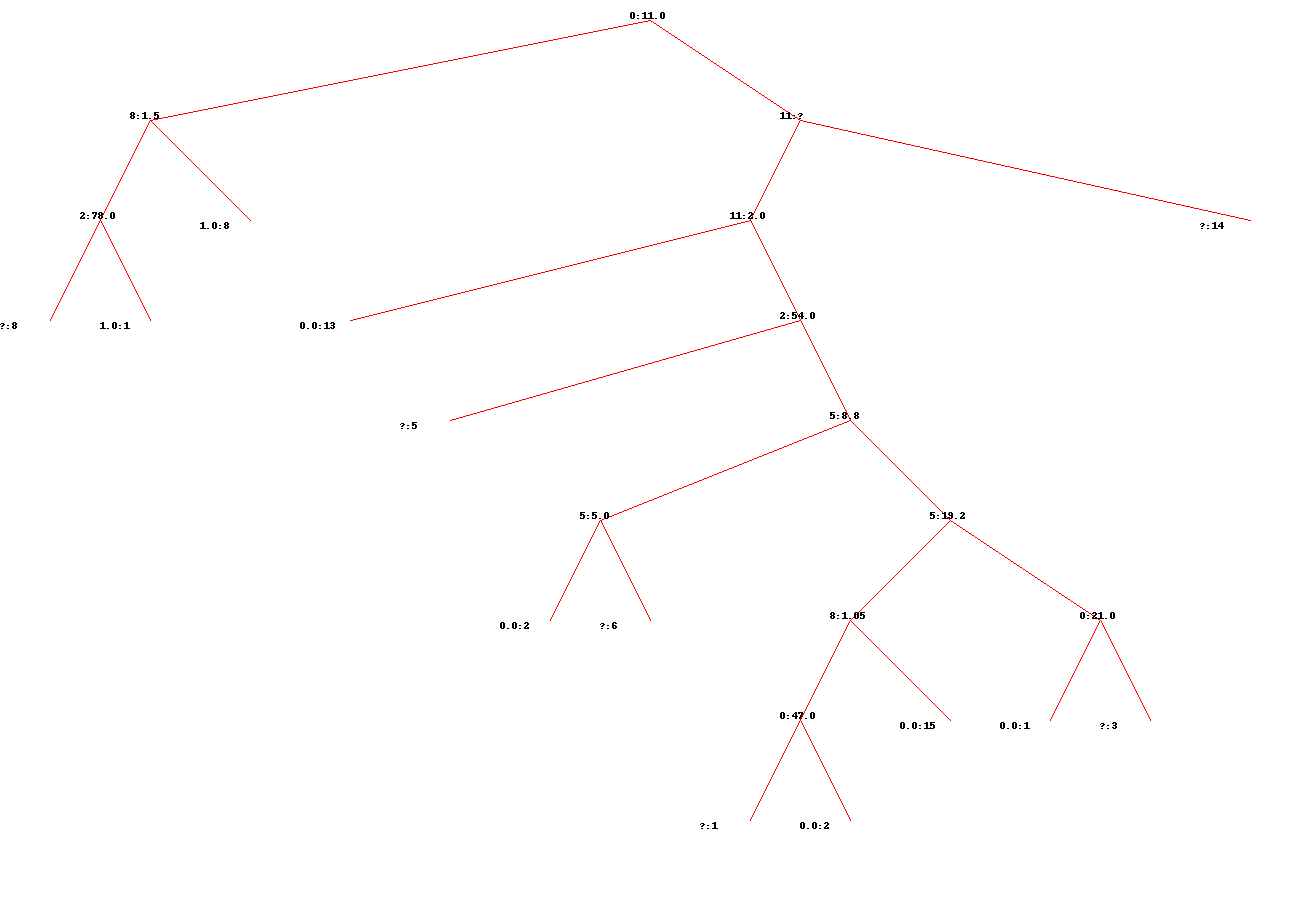
**Figura 5:** Albero training\_set = 40% e test\_set = 60%

****

**Echocardiogram training\_set=60%, test\_set=40%**

Usando il 60% dati per il training (ovvero 79 istanze), abbiamo che l’albero (**Figura 7**) ha profondità massima 8, quindi incomincia a diventare meno generale, ed un grafico delle performance tendenzialmente crescente con un massimo di 0.89, ma con alcune fasi di decrescita nel passaggio tra il 30%-40% e 60%-80% del training set size. (**Figura 6**)

**Figura 6:** Performance training\_set = 60% e test\_set = 40%

****

**Figura 7:** Albero training\_set = 60% e test\_set = 40%

Si nota dai grafici delle performance che la learning curve sale molto lentamente, questo perché abbiamo usato una rappresentazione del mondo ridondante ed è possibile migliorare la stima eliminando dalla tabella della rappresentazione del mondo di Echocardiogram degli attributi.

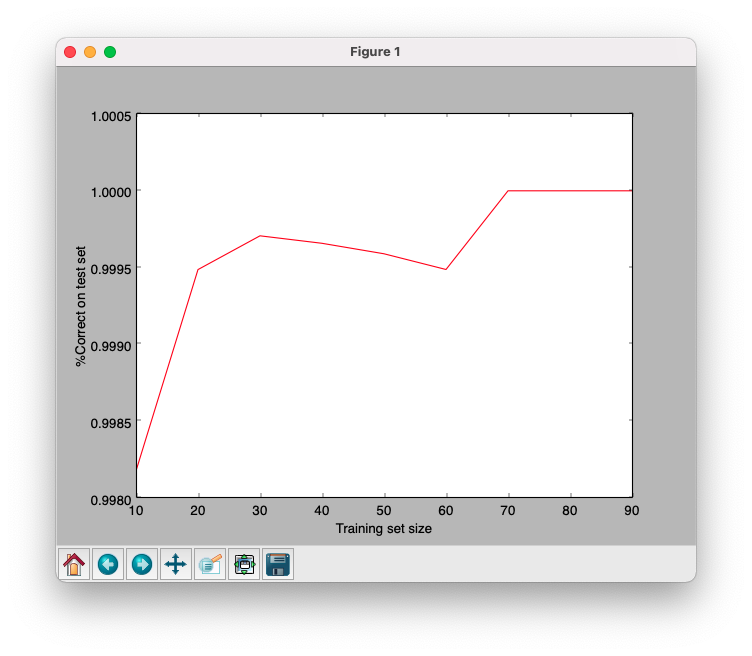
**1.3 Dataset Mushrooms**

**Scopo:** Nel dataset sono presenti i campioni di 23 specie di funghi con branchie nella famiglia Agaricus e Lepiota. Ogni specie è identificata come sicuramente commestibile, decisamente velenosa, o di commestibilità sconosciuta e sconsigliata. Quest'ultima classe è stata combinata con quella velenosa. Quindi l’albero decisionale deve identificare quali funghi sono commestibili o meno.

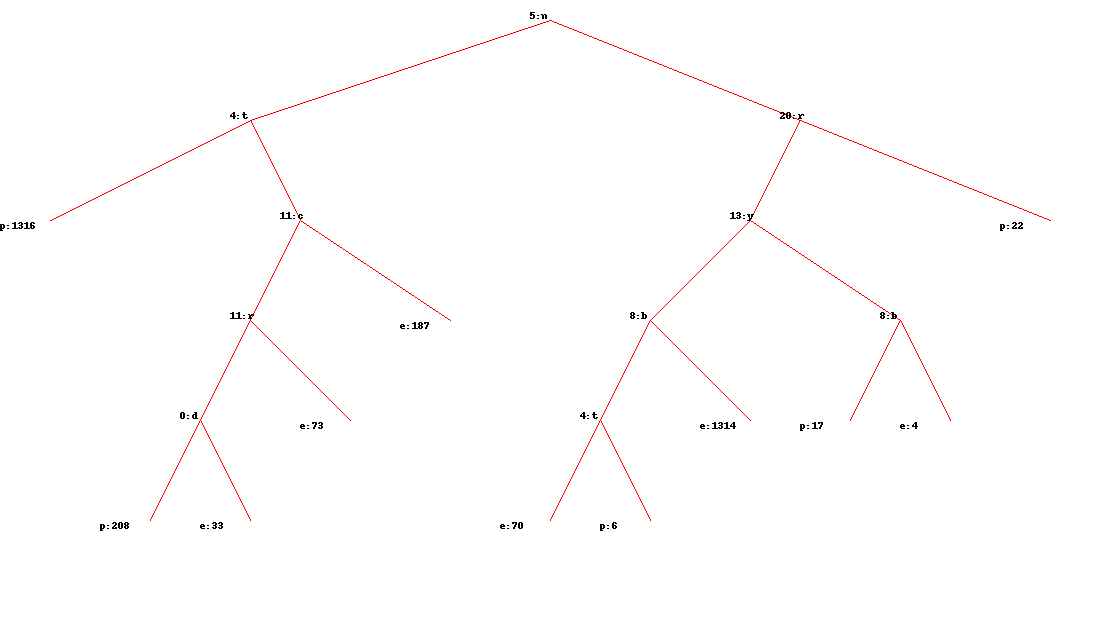
Il dataset è caratterizzato da ben 21 attributi e 8124 istanze.

**Mushrooms training\_set=40%, test\_set=60%**

Usando il 40% dati per il training (ovvero 3250 istanze), abbiamo che l’albero (**Figura 8**) ha profondità massima 6, ed un grafico delle performance tendenzialmente crescente fino ad un massimo di circa 1.0 con alcuni punti in corrispondenza del 28%-60%, in cui la performance si riduce. (**Figura 9**)

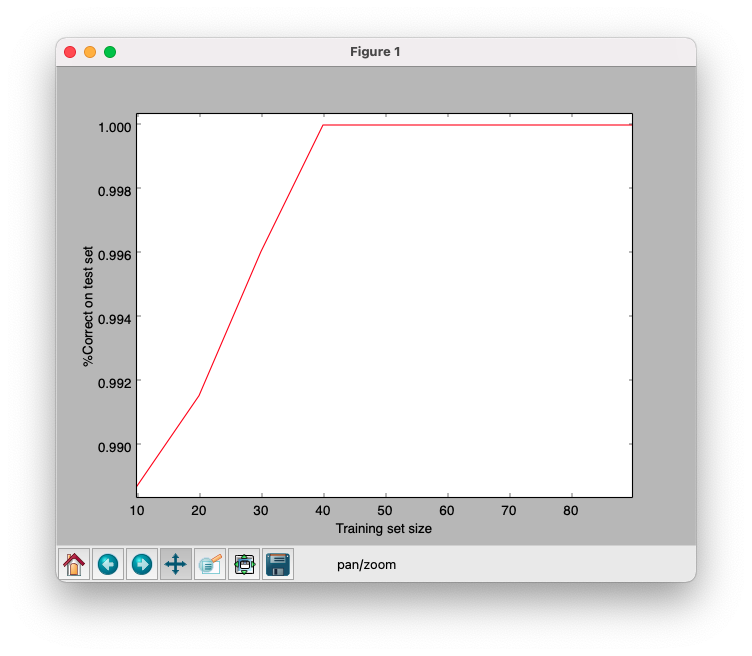
****

**Figura 8:** Performance training\_set = 40% e test\_set = 60%

****

**Figura 9:** Albero training\_set = 40% e test\_set = 60%

**Mushrooms training\_set= 60%, test\_set=40%**

****

Usando il 60% dati per il training (ovvero 4874 istanze), abbiamo che l’albero (**Figura 11**) ha profondità massima 6, ed un grafico delle performance tendenzialmente crescente fino ad un massimo di circa 1.0. (**Figura 10**)

**Figura 10:** Performance training\_set = 60% e test\_set = 40%

**Immagine che contiene cielo, testo, mappa

Descrizione generata automaticamente**

**Figura 11:** Albero training\_set = 60% e test\_set = 40%

Si nota dai grafici delle performance che la learning curve si avvicina molto alla curva realizzabile (dalla **Figura 3**) dunque abbiamo appreso il modello perfetto.

**ESERCIZIO 2**

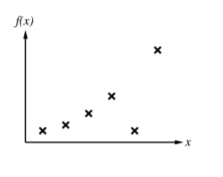
**2.1 Perché' un agente in grado di apprendere Alberi di Decisione è considerato Intelligente?**

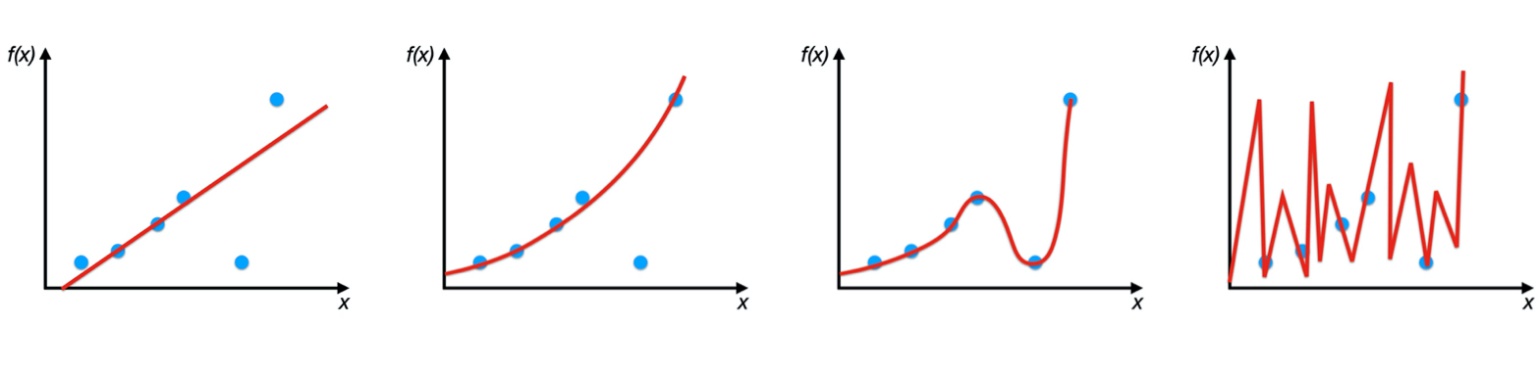
Un agente in grado di apprendere un albero di decisione è definito intelligente perché esso attraverso un processo ***induttivo****,* cioè a partire dalla descrizione di esempi di comportamento, di approssimare la ***funzione target f***, che rappresenta la funzione che l’agente deve apprendere.

In particolare, l’agente classifica gli esempi (**positivi e negativi**) in base a delle euristiche proprie del problema di interesse, maggiori saranno gli esempi e più *l’ipotesi h* approssimerà in maniera migliore la f e quindi, spiegherà con maggior correttezza gli esempi dati. Tuttavia, non si limita a trovare una qualsiasi ***h***, ma cerca una ***h*** che in base al “rasoio di Ockham” deve essere la più semplice possibile per generalizzare meglio il problema senza generare il fenomeno dell’***overfitting***, cioè quel fenomeno in cui il modello spiega così bene gli esempi forniti in fase di test che non riesce a spiegare gli esempi futuri.

**Esempio: Curve fitting**

Per comprendere meglio il fenomeno dell’**overfitting**, consideriamo gli esempi rappresentati nella **Figura 4** dalle **X** del grafico. Se si cerca una funzione che spiega quei punti (passa per tali punti), si può notare che presa una qualsiasi funzione coerente con i dati (retta, esponenziale, sinusoide, spezzata) all’aumentare degli esempi spiegati aumenta la complessità della funzione da studiare, e dunque, risulta più conveniente scegliere la retta di regressione rossa piuttosto che la funzione arancione o quella verde.

****



**Figura 12**: Rappresentazione grafica di un insieme di esempi

**Figura 13**: Diverse funzioni che interpolano i punti

**2.2 Cosa si intende con il termine apprendimento quando si utilizza il DT learning?**

Con il termine apprendimento nel contesto del DT learning intendiamo quel processo attraverso il quale, data la ***“Tabella dei sistemi di apprendimenti”*** composta sulle righe dagli esempi dati all’agente, ogni esempio avrà un insieme di colonne che caratterizzano le proprietà del sistema ed infine, una colonna target che caratterizza la classificazione (**vero o falsa**) dell’esempio che si vuole apprendere. A partire da tale tabella si costruisce attraverso un algoritmo di ***DTL (Decision Tree Learning)*** un albero che permette ad esempio, di rappresentare in maniera automatica il comportamento di un essere umano intelligente, che si trova dinanzi al dover effettuare una scelta.

L'algoritmo elabora questi dati in un modo non previsto dal programmatore. L'essere umano non sa a priori come rappresentare il mondo e quali attributi mettere nella tabella. Si procede per raffinamenti successivi, vedendo se il risultato è soddisfacente applicando la tecnica di intelligenza artificiale.

Un albero è un ***direct acyclic graph (DAG)***. Dalla radice si diramano i rami dell'albero. Scendiamo fino ad arrivare alle foglie. L'albero di decisione è fatto in questo modo: dentro alle scatole intermedie ci sono delle condizioni. Quando si arriva alla foglia, si giunge alla classificazione. Le condizioni sono i nomi degli attributi e da ogni nodo escono tanti archi quanti sono i valori che l'attributo può prendere.

La differenza tra l'informatica tradizionale e l'informatica che si dedica all'intelligenza artificiale è proprio il concetto di learning in cui l’agente apprende e non esegue un codice statico, scritto precedentemente, ma impara dalle sue interazioni con l'ambiente. Tramite uno score è possibile parametrare le performance e far capire all’agente se si sta comportando in modo positivo aumentando lo score o se si sta comportando in modo negativo penalizzando lo score.

**ESERCIZIO 3**

**3.1 Spiegare l’algoritmo A\***

L’algoritmo A\* fa parte della famiglia degli algoritmi Best-First Search, cioè degli algoritmi di ricerca informata (basati su grafi) che si basano sulle informazioni date sul problema per scegliere la soluzione nello spazio delle soluzioni candidate; ovvero la sequenza di passi che porta da uno stato iniziale ad uno stato obiettivo. E quindi, tale algoritmo si adopera quando siamo interessati non solo alla sequenza di passi, ma anche al costo associato al cammino, in particolare si associa ad ogni nodo un grado di desiderabilità (in genere, quello più desiderabile è quello che si avvicina maggiormente alla soluzione). Il criterio usato è dato dalla funzione:

in cui:

* g(n) è il costo storico per arrivare a partire dal nodo radice al nodo n-esimo
* h(n) è un’euristica che ci permette di stimare il costo che impieghiamo per arrivare al nodo obiettivo a partire dal nodo n.

**3.2 Descrivere i casi in cui si applica e cosa si intende per euristica ammissibile in A\***

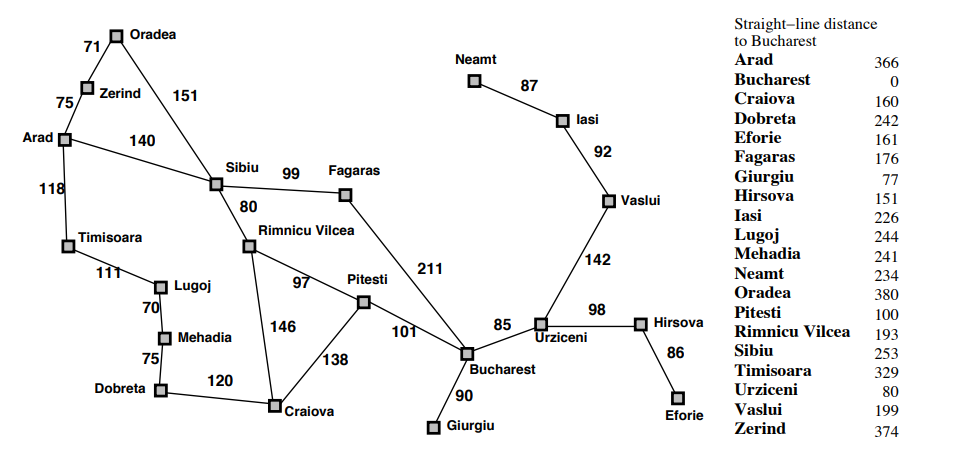
L’algoritmo A\* risulta efficace nelle situazioni in cui cerchiamo all’interno di un albero di ricerca la soluzione ottimale e completa. Ciò avviene solo e soltanto se ci ritroviamo in uno spazio di stati limitato e se utilizziamo un’euristica h(n) ammissibile. L’euristica influenza fortemente i risultati conseguiti da A\*. Esso, in particolare, ne determina il tempo complessivo di esecuzione e la qualità della soluzione finale.

La funzione euristica si dice ammissibile se il valore da essa stimato è minore o uguale al costo reale per raggiungere la soluzione a partire dal nodo che stiamo considerando.

Dunque, un’euristica è ammissibile quando l’errore di stima non è mai in eccesso.

**3.3 Mostrare un esempio di applicazione di A\* su un problema di navigazione stradale**

**Si utilizza la mappa della Romania per applicare l’algoritmo A\* nella risoluzione del problema di navigazione stradale tra Oradea e Bucharest.**



Si rappresenta tramite un grafo la mappa della Romania

Si rappresenta il problema creando un'astrazione della mappa della Romania con l’utilizzo di un grafo per far sì che sia facilmente rappresentabile tramite l’algoritmo A\*.  La rappresentazione a grafo sarà una sequenza di archi che porta dallo stato iniziale ad uno stato obiettivo. L’insieme delle possibili soluzioni comprende tutti i possibili percorsi, giusti o sbagliati sul grafo (senza vincoli sul punto di partenza e sul punto di arrivo).

La scelta dell’euristica ammissibile

Si tratta di un criterio approssimativo per trovare una soluzione “accettabile” per il nostro problema.

Nel nostro caso utilizziamo l’euristica della **distanza a linea d’aria**di ogni città della Romania verso Bucharest, indicato nella tabella a destra della mappa. Con una buona funzione euristica la complessità può essere ridotta considerevolmente: l’entità precisa dipende dalla natura del problema e dalla quantità dell’euristica.

**Sviluppo dell’albero di Ricerca tramite Algoritmo A\***

**STATO INIZIALE**

*380+0*

Lo stato iniziale rappresentato dalla radice del grafo è ***Oradea***, il nostro punto di partenza. La funzione f(n) sarà uguale a 380 dato che è presente solo la componente h(n) indicante la distanza in linea d’aria da Bucharest, mentre il costo storico g(n) è zero essendo il nodo radice.

**ESPANSIONE DI ORADEA**

*71+374 = 485*

***151+253 = 404***

All’espansione di Oradea, la funzione f(n) sarà costruita per i corrispettivi sottoalberi. Il prossimo nodo da espandere sarà **Sibiu** dato che ha f(n) minore rispetto a quella di Zerind. Infatti, avremo che la f(n) sarà data dalla somma di 151 che rappresenta la distanza Oradea-Sibiu percorsa e quindi il costo storico, mentre 253 è la distanza in linea d’aria Sibiu-Bucharest.

**DOPO ESPANSIONE DI SIBIU**

*71+374 = 485*

*151+99+176 = 426*

***151+80+193 = 424***

*151+140+366 = 657*

*151+151+380 = 682*

All’espansione di Sibiu si procede in modo analogo al passo precedente costruendo la funzione f(n) per i rispettivi sottoalberi. La scelta questa volta ricade su **Rimnicu Vilcea** dato che ha f(n) minore.

**DOPO ESPANSIONE DI RIMNICU VILCEA**

*71+374 = 485*

*151+140+366 = 657*

***151+99+176 = 426***

*151+151+380 = 682*

*151+80+97+100 = 428*

*151+80+146+100 = 537*

*151+80+80+253 = 564*

Con l’espansione di Rimnicu Vilcea, saremmo portati ad espandere Pitesti, questo però si rivela un vicolo cieco. Infatti, applicando l’algoritmo A\* esso ci dice che il nodo precedente Fagaras (appartenente alla frontiera), ha f(n) minore rispetto a questo. L’algoritmo allora decide di tornare indietro e espandere Fagaras.

**DOPO ESPANSIONE DI FAGARAS**

*71+374 = 485*

*151+140+366 = 657*

*151+151+380 = 682*

*151+99+211+0 = 461*

*151+80+80+253 = 564*

***151+80+97+100 = 428***

*151+80+146+100 = 537*

*151+99+99+253 = 502*

A questo punto con l’espansione di Fagaras e avendo ottenuto la funzione costruita per i sottoalberi, otteniamo il nodo Bucharest ma l’algoritmo A\* ci dice che la soluzione non è ottimale, perché il nodo Pitesti ha una f(n) minore, allora decidiamo di espandere Pitesti nell’altro sottoalbero.

**DOPO ESPANSIONE DI PITESTI**

*71+374 = 485*

*151+140+366 = 657*

*151+151+380 = 682*

*151+99+99+253 = 502*

*151+80+80+253 = 564*

*151+99+211+0 = 461*

*151+80+146+100 = 537*

*151+80+97+138+160 = 626*

*151+80+97+97+153 = 618*

***151+80+97+101+0 = 429***

Arrivati qui, all’espansione di Pitesti otteniamo nuovamente Bucharest,questa volta l’A\* ci dice che questa è una soluzione ottimale, perché il nodo Bucharest trovato ha la f(n) minore tra tutti i nodi nella frontiera.

**ESERCIZIO 4**

**4.1 Descrivere con un esempio come si utilizza un albero di decisione**

Gli alberi di decisione (o alberi decisionali) sono una rappresentazione dell'apprendimento nel machine learning. Un albero di decisione è un sistema con ***n*** variabili in input e***m*** variabili in output. Le variabili in input (attributi) sono derivate dall'osservazione dell'ambiente. Le ultime variabili in output, invece, identificano la decisione / azione da intraprendere.

Il processo decisionale è rappresentato con un albero logico rovesciato dove ogni nodo è una funzione condizionale.

Ogni nodo verifica una condizione (test) su una particolare proprietà dell'ambiente (variabile) e ha due o più diramazioni verso il basso in funzione della condizione. Il processo consiste in una sequenza di test. Comincia sempre dal nodo radice, il nodo genitore situato più in alto nella struttura, e procede verso il basso.

A seconda dei valori rilevati in ciascun nodo, il flusso prende una direzione oppure un’altra e procede progressivamente verso il basso.

La decisione finale si trova nei nodi foglia terminali, quelli più in basso. In questo modo, dopo aver analizzato le varie condizioni, l'agente giunge alla decisione finale.

Per la realizzazione dell’albero segue il concetto di scelta del migliore attributo.

**L’Algoritmo DTL**

**funzione** DTL (esempi, attributi, default) **ritorna** un albero

*SE* non ci sono esempi allora **ritorna** default

*SE* tutti gli esempi hanno la stessa classificazione **ritorna** la classificazione

*SE* non ci sono attributi **ritorna** la Moda(esempi)

*ALTRIMENTI*

Sia best ← Choose-Attribute (attributi, esempi)

Sia albero ← un nuovo albero di decisione con radice di test best

*PER OGNI* valore vi di best *FAI*

esempii ← {elementi di esempi con best = vi}

sotto-albero ←DTL (esempii, attributi−best, Mode(esempi))

*AGGIUNGI* un ramo ad albero con etichetta (best= vi) e sottoalbero sotto-albero

**RITORNA** albero

Questa funzione è ricorsiva. Prende in ingresso esempi, attributi e una classificazione di default. Gli esempi sono le righe della tabella dei sistemi di apprendimento mentre gli attributi sono le colonne.

• Se l’insieme degli esempi è vuoto, allora restituisce la classificazione di default.

• Se tutti gli esempi hanno la stessa classificazione, allora restituisce la classificazione stessa.

• Se l’insieme degli attributi è vuoto, allora restituisce la moda degli esempi, che conta la classe degli esempi e ritorna la classe che occorre maggiormente.

• Negli altri casi, scegli un attributo, crea un nuovo albero di decisione avente come radice l’attributo scelto, e per ogni valore che l’attributo può assumere, riduci l’insieme degli esempi, creando un sottoinsieme che contenga solo gli esempi per cui l’attributo vale l’i-esimo valore.

In questo modo, sto partizionando l’insieme degli esempi sulla base del valore assunto dall’attributo. Si crea un sottoalbero, passando alla funzione DTL, chiamata per ricorsione, il sottoinsieme degli esempi precedentemente determinato, l’insieme degli attributi eccetto quello scelto, e la moda degli esempi come classificazione di default. Infine, il sottoalbero viene collegato all’albero principale.

Questo risulta essere un meta algoritmo, in quanto non siamo in grado di specificare tutti i dettagli, ma solo la struttura dell’algoritmo.

Analizziamo il problema di dover decidere se attendere al di fuori di un ristorante oppure cambiare locale, per risolvere questo problema utilizzeremo la tabella degli esempi e realizzeremo l’albero. (Sarà spiegato al meglio nel prossimo esercizio).

Prendendo in prestito l’albero già costruito abbiamo una schematizzazione ripartitiva che ci permette di stabilire una regola generale per decidere appunto se attendere oppure no.

**ESERCIZIO 5**

**5.1 Descrivere con un esempio il processo di apprendimento di un albero di decisione**

Si prova a costruire un albero per decidere se aspettare che si liberi un tavolo al ristorante oppure no.

Per fare ciò come detto precedentemente abbiamo bisogno di una “Tabella dei sistemi di apprendimenti” dove andremo ad inserire i vari esempi su cui ci baseremo per la realizzazione dell’albero.

Nel nostro caso il target della tabella sarà l’attendere o meno un tavolo.

|  |  |  |
| --- | --- | --- |
| ATTRIBUTI | | TARGET |
| Type: Qual è il genere di cucina? | *French, Italian, Thai, Burger* | WillWait |
| Friday/Sat: È il fine settimana? | *Yes, No* |
| Hungry: Si è affamati? | *Yes, No* |
| Price: Il ristorante è costoso? | *$$$, $$, $* |

Un albero di decisione booleano è una coppia (X, y) dove:

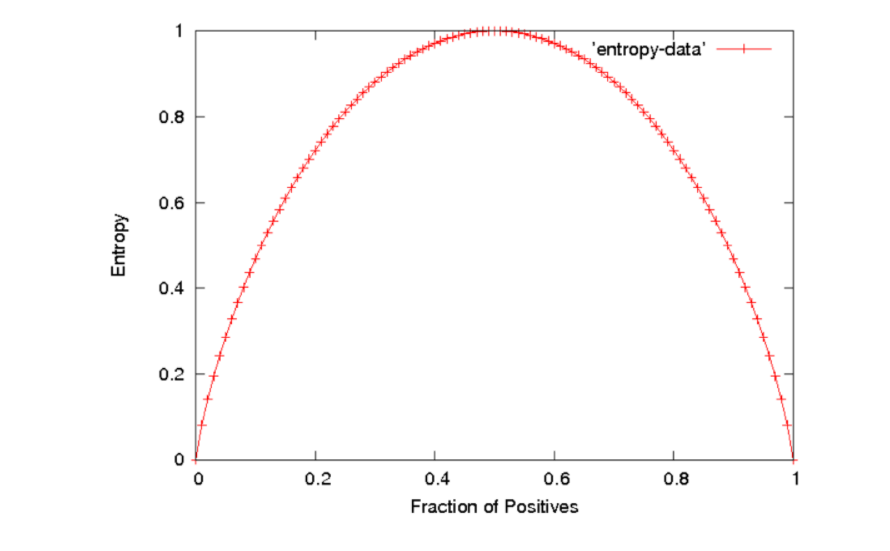
* X è un vettore di valori per gli attributi
* y ∈ {True, false}

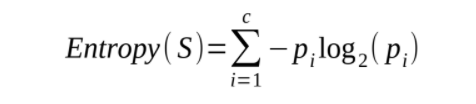
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| EX | Type | Friday/Sat | Hun | Price | WillWait |
| x1 | French | False | True | $$$ | y1=True |
| x2 | Thai | False | True | $ | y2=False |
| x3 | Burger | False | False | $ | y3=True |
| x4 | Thai | True | True | $ | y4=True |
| x5 | French | True | False | $$$ | y5=False |
| x6 | Italian | False | True | $$ | y6=True |

***Quale attributo utilizziamo prima per effettuare la decisione?***

Per trovare l’attributo più discriminante veniamo in contro alla funzione **entropia**, che ci permette di calcolare il grado di disordine all’interno di un sistema e quindi il migliore attributo da espandere.

La funzione di entropia è una rappresentazione binaria e può assumere valori da 0 a 1 ed ha una forma:





**Figura 14**: Funzione di entropia per classificazione binaria

**CALCOLO DELL’ENTROPIA DEI SINGOLI ATTRIBUTI**

Il calcolo si suddivide in due logaritmi in base due; il primo individua i **false**, il secondo individua i **true**.

Dunque, il nodo radice sarà

French

Thai

**???**

**True**

Burger

Italian

**???**

**True**

Si continuano i calcoli estrapolando il sottoinsieme **TYPE = THAI** degli attributi:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| EX | Friday/Sat | Hun | Price | WillWait |
| x2 | False | True | $ | y2=False |
| x4 | True | True | $ | y4=True |

I calcoli saranno:

Dunque, l’albero al passo successivo sarà il seguente:

**???**

**True**

Italian

**True**

**???**

French

Thai

Burger

False

True

**???**

Si continuano i calcoli estrapolando il sottoinsieme **TYPE = FRENCH** degli attributi:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| EX | Friday/Sat | Hun | Price | WillWait |
| x1 | False | True | $$$ | T |
| x5 | True | False | $$$ | F |

Dunque, l’albero al passo successivo sarà il seguente:

**???**

**True**

Italian

**True**

**???**

**???**

French

**???**

Thai

Burger

False

True

True

False

Seguono gli altri sottoinsiemi:

Si estrapola il sottoinsieme **FRI/SAT=FALSE** e restituisce **FALSE**

|  |  |  |  |
| --- | --- | --- | --- |
| EX | Hun | Price | WillWait |
| x2 | True | $ | y2=False |

Si estrapola il sottoinsieme **FRI/SAT=TRUE** e restituisce **TRUE**

|  |  |  |  |
| --- | --- | --- | --- |
| EX | Hun | Price | WillWait |
| x4 | True | $ | y4=True |

Si estrapola il sottoinsieme **HUN=TRUE** e restituisce **TRUE**

|  |  |  |  |
| --- | --- | --- | --- |
| EX | Friday/Sat | Price | WillWait |
| x1 | False | $$$ | y1=True |

Si estrapola il sottoinsieme **HUN=FALSE** e restituisce **FALSE**

|  |  |  |  |
| --- | --- | --- | --- |
| EX | Friday/Sat | Price | WillWait |
| X5 | True | $$$ | y5 =False |

In definitiva l’albero di decisione per la tabella di apprendimento fornita è questo:

**True**

**True**

Italian

**True**

**True**

**False**

French

**False**

Thai

Burger

True

False

True

False

**ESERCIZIO 6**

### **6.1 Cosa sono: lo spazio degli stati, il grafo degli stati e l’albero di ricerca (Tree Search)**

Un primo esempio di astrazione del problema da risolvere è un **grafo degli stati,**ovvero un tipo di diagramma orientato che descrive il comportamento di un problema attraverso nodi (stato) e frecce (collegamenti).

L’obiettivo del Problem Solving è la ricerca dello spazio degli stati il cui risultato è una sequenza di azioni che intercorre tra lo stato iniziale a quello finale (***soluzione del problema***).

**Lo spazio degli stati**è l’insieme di tutti gli stati che sono raggiungibili dallo stato iniziale attraverso sequenza di azioni (cammino) determinati dall’algoritmo di ricerca (**albero di ricerca**).

Il **Tree Search** è la rappresentazione delle sequenze di azioni che un determinato agente effettua a partire da uno stato iniziale (il nodo padre). Dal nodo padre sono collegati gli stati conseguenti per ogni azione possibile. Ogni azione possibile viene rappresentata tramite un collegamento che unisce il nodo padre ed il nodo figlio (i nodi dell’albero di ricerca sono gli stessi dello spazio degli stati). Dal nodo figlio possono espandersi nuovi nodi sulla base dell’esigenza del problema (espansione) ed un nodo senza ulteriori azioni possibili è detto nodo foglia e contribuisce la frontiera degli stati.

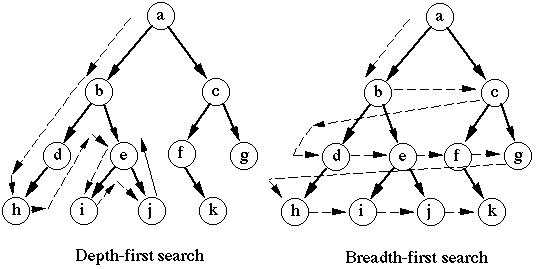
### **6.2 Descrivere in modo sintetico il meta algoritmo generale di costruzione di un Tree Search**

Per costruire un Tree Search, si segue un algoritmo al quale vengono forniti in input un ***problema*** ed una ***frontiera***(ovvero l’insieme dei nodi da espandere) e che restituisce in output l’albero di ricerca. Si inizia inserendo nella frontiera il nodo iniziale del problema, poi si esegue un ciclo infinito in cui si verifica se la frontiera è vuota, se lo è ritorna un fallimento. Altrimenti preleva un nodo dalla frontiera e testa se è un nodo obiettivo del problema e quindi, se lo è ritorna il nodo in questione. Se il nodo non è obiettivo, si espande tale nodo con una funzione ***EXPAND()*** e lo si inserisce nella frontiera, il ciclo può così iniziare da capo. La funzione ***EXPAND()***ritorna un insieme di nodi, si inizializza un insieme vuoto in cui andranno inseriti i ***nodi “successori”*** di quello fornito in input, inoltre, viene fornito anche il problema alla funzione. Si esegue un ciclo per ciascun stato raggiungibile da quello attuale determinato dalla ***funzione successore S(x)***, si crea un nodo per quello stato, si aggiungono a tale nodo tutte le informazioni che lo caratterizzano (nodo padre, azione, ovviamente lo stato che individua, il costo di cammino e la profondità). Si aggiunge tale nodo all’insieme dei nodi “successori”, si verifica se ci sono altri stati raggiungibili e in caso affermativo si ripete il ciclo, altrimenti si ritorna alla funzione di costruzione del Tree Search l’insieme dei nodi espansi.

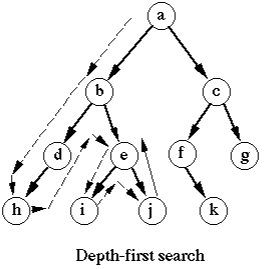
### **6.3 Spiegare in modo sintetico le caratteristiche degli algoritmi Tree Search visti a lezione**

I problemi di intelligenza artificiale che riguardano il problem solving vengono sempre ricondotti a problemi di ricerca su alberi. Gli algoritmi di ricerca si dividono in due categorie, quelli ***non informati*** (Breadth-first search, Depth-first Search e Uniform-cost search) e quelli ***informati*** (Best-first search, che si specializza in due varianti Greedy Search e A\* Search).

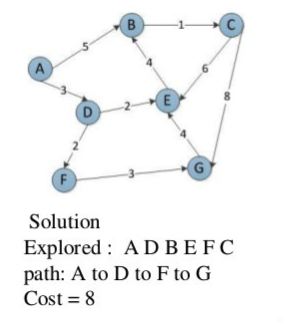
La prima categoria di algoritmi, come dice il nome, non sfruttano nessuna informazione sul numero di passi o sul costo di cammino dallo stato attuale a quello obiettivo, iniziamo con il ***Breadth-first search*** che effettua una ricerca in ampiezza, i nodi della frontiera vengono inseriti in una ***coda gestita tramite politica FSEO***. Prendendo come esempio l’albero in figura, all’inizio nella coda viene inserito il nodo a, poi questo nodo viene prelevato dalla coda ed espanso, poi vengono aggiunti i nodi b e c, essendo che b viene inserito per primo nella coda sarà anche il primo ad essere espanso tra i due, e così via. Quest'algoritmo è completo (perché trova una soluzione se esiste), è ottimale se il costo di cammino è 1 per ogni step perché la soluzione trovata è quella a profondità minore, tuttavia in generale non è ottimale.



Il secondo algoritmo della categoria è il ***Depth-first search***, esso memorizza i nodi della frontiera in uno ***stack gestito con politica LSEO***, questa ricerca viene anche detta in profondità, usando sempre l’esempio in figura abbiamo che il nodo a viene espanso allora nello stack vengono inseriti nell’ordine c e b, a questo punto l’ultimo nodo inserito è b. Viene prelevato b, quindi viene espanso e vengono inseriti nello stack i nodi e ed d, e così via. Quest'algoritmo non è né completo né ottimale, perché se ci sono dei cicli nell’albero esso lo percorre all’infinito, inoltre non trova la soluzione migliore, ma la prima soluzione che riesce a calcolare.



L’ultimo algoritmo tra quelli non informati è il ***Uniform-cost search***, che dispone i nodi della frontiera all’interno di una ***coda ordinata in maniera crescente sulla base del costo di raggiungimento di un nodo***, anche se appartiene agli algoritmi non informati, in realtà porta con sé un'informazione maggiore rispetto ai precedenti rappresentata dal costo di cammino. Un esempio di applicazione è mostrato in figura, dove si nota che il percorso di ricerca su quest’albero segue i percorsi con costo minore.



Passiamo all’altra categoria di algoritmi di ricerca, ovvero quelli informati, per capire questi algoritmi dobbiamo introdurre il concetto di euristica, ovvero una regola empirica che ci permettono di trovare una buona soluzione al problema senza dimostrarlo analiticamente. Tra gli algoritmi informati, quello più importante è il ***Best-first search***, i nodi della frontiera vengono ordinati secondo un ***criterio di desiderabilità***(quello più desiderabile è quello che più si avvicina alla soluzione), questa è l’euristica su cui si basa l’algoritmo, che poi può essere implementata con diverse strategie, tra cui il Greedy Search e A\* Search.

Per implementare il ***Greedy Search (ricerca “avida”)*** è necessaria una funzione di desiderabilità, che può essere ad esempio, la distanza in linea d’aria con una città obiettivo nel caso della navigazione stradale. Quindi la scelta del nodo della frontiera da espandere viene fatta sulla base del valore di questa euristica h(n) prendendo quella con il valore minore. Quest’algoritmo è completo, se gestisco i cicli all’interno dell’albero, ma non è ottimale, tuttavia, trova una soluzione sub-ottimale (massimo locale).

L'algoritmo ***A\* Search*** si basa non solo su un’euristica ma anche su un costo storico, infatti la funzione di desiderabilità è ottenuta in questo modo: f(n) = g(n) + h(n), in cui g(n) è il costo passato per raggiungere il nodo n a partire dal nodo iniziale, mentre h(n) è la stima di quanto sia distante n dallo stato obiettivo. Quest’algoritmo è ottimale se h(n) è ammissibile cioè il costo stimato è minore o uguale a quello reale per raggiungere l’obiettivo a partire da n, mentre è completo se ci sono un numero finito di stati.