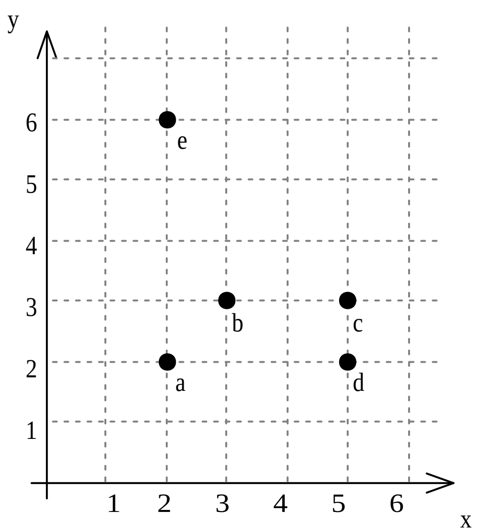
Apprendimento Automatico
Secondo Esonero A.A. 2007/2008
Prof. Roberto Navigli

1 Apprendimento Non Supervisionato

Esercizio 1.1 (5 punti). Determinare il clustering gerarchico agglomerativo delle seguenti istanze sul piano bidimensionale utilizzando la strategia single link.



Quesito 1.2 (3 punti). Qual è la differenza tra apprendimento supervisionato e non supervisionato?

2 Significatività Statistica

Esercizio 2.1 (5 punti). Due classificatori c_1 e c_2 sono stati applicati a un problema di apprendimento. L'accuratezza dei due sistemi, misurata sul medesimo test set T costituito da 1000 istanze, è riportata nella tabella seguente. Determinare se la differenza di prestazioni tra c_1 e c_2 è statisticamente significativa ($p < 0.05$, valore critico di χ^2 per un grado di libertà pari a $3,84$).

classif.	**corretti**	**errati**	**accuratezza**
c_1	820	180	82%
c_2	780	220	78%

3 Algoritmi Genetici

Esercizio 3.1. In un videogioco, un livello di gioco è rappresentato da un insieme di stanze tra loro collegate. Un livello è un grafo diretto rappresentato mediante una matrice di adiacenza A, in cui $a_{ij} = 1$ se dalla stanza i è possibile spostarsi nella stanza j, 0 altrimenti. Si vuole applicare un algoritmo genetico per generare un livello in cui, per ogni coppia di stanze i e j, esiste almeno un cammino da i a j. In altre parole, si vuole apprendere un grafo diretto connesso.

a) Fornire una strategia di crossover double point per l'algoritmo genetico **(2 punti)**;

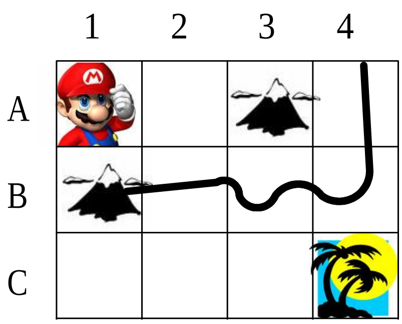
b) Definire una funzione di fitness ragionevole, che tenga anche conto di quanto il numero di archi sia vicino a un numero k desiderato (notare però che possono esistere grafi non connessi aventi k archi) **(3 punti)**;

c) Applicare un passo dell'algoritmo genetico alla seguente popolazione iniziale, con parametri $r = 2/3$ e $m = 2/3$ **(3 punti)**:

$$P = \left\{ \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \right\}$$

4 Apprendimento con Rinforzo

Esercizio 4.1 (6 punti). Mario deve determinare la migliore politica per raggiungere la palma. Applicare la prima iterazione dell'algoritmo Q-Learning al seguente ambiente supponendo che Mario si sposti a partire dalla cella in cui si trova seguendo il percorso meno impervio. Calcolare tutti i valori di $\hat{Q}$ (procedendo anche a ritroso nel percorso) per $\gamma = 0.9$. Le ricompense per le azioni di spostamento (orizzontale e verticale) sono le seguenti: 0 se ci si sposta in una cella vuota, -10 in una cella di fiume, -100 in una cella di montagna, +10 quando si giunge alla palma.



5 Naive Bayes

Esercizio 5.1 (6 punti). Si supponga di rappresentare le strade di una regione mediante un vettore di 3 attributi: (asfalto, numero incroci, limite velocità). Dato il seguente insieme di addestramento, si vuole determinare mediante Naive Bayes la pericolosità di una strada $x =$ (buono, molti, 90).

asfalto	**incroci**	**limite**	**pericolosa**
buche	*nessuno*	70	*no*
sgretolato	*molti*	90	*sì*
buono	*pochi*	90	*no*
buche	*molti*	90	*sì*
buche	*pochi*	70	*sì*
buono	*molti*	110	*sì*
sgretolato	*molti*	50	*no*

Quesito 5.2 (3 punti). Discutere brevemente i vantaggi dell'apprendimento probabilistico rispetto ad altre tecniche di apprendimento automatico.

6 Soluzioni

Esercizio 1.1. Il clustering iniziale $c^{(0)} = \{\{a\}, \{b\}, \{c\}, \{d\}, \{e\}\}$.

$d(a, b) = \sqrt{1^2 + 1^2} = \sqrt{2}$
$d(a, c) = \sqrt{3^2 + 1^2} = \sqrt{10}$
$d(a, d) = 3$
$d(a, e) = 4$
$d(b, c) = 2$
$d(b, d) = \sqrt{2^2 + 1^2} = \sqrt{5}$
$d(b, e) = \sqrt{1^2 + 3^2} = \sqrt{10}$
$d(c, d) = 1$
$d(c, e) = \sqrt{3^2 + 3^2} = \sqrt{18} = 3\sqrt{2}$
$d(d, e) = \sqrt{3^2 + 4^2} = \sqrt{25} = 5$

La coppia di cluster che massimizza la similarità (ovvero minimizza la distanza) tra gli elementi è $(c,\ d)$. I due cluster vengono fusi in un unico cluster: $\{c, d\}$, ottenendo il clustering $c^{(1)} = \{\{a\}, \{b\}, \{c, d\}, \{e\}\}$. Alla seconda iterazione, la coppia di cluster più simile è $(a,\ b)$. Otteniamo quindi il clustering $c^{(2)} = \{\{a, b\}, \{c, d\}, \{e\}\}$. Alla terza iterazione, la coppia di cluster più simile è $(\{a, b\}, \{c, d\})$, grazie alla distanza minima $d(b, c)$. Otteniamo il clustering $c^{(3)} = \{\{a, b, c, d\}, \{e\}\}$. Infine, il clustering finale è dato dall'unione dei due cluster rimanenti.

Quesito 1.2. I metodi di apprendimento supervisionato apprendono un modello del problema a partire da un insieme di dati di addestramento. In tali metodi, l'insieme delle classi è noto a priori e ciascun esempio nell'insieme di dati è annotato con la classe corretta. Al contrario, i metodi non supervisionati apprendono a partire da insiemi di dati non annotati e le classi da associato agli esempi non sono note a priori.

Esercizio 2.1. La tabella di contingenza, aggiornata con il conteggio medio, è la seguente:

classif.	**corretto**	**errato**
c_1	820 (800)	180 (200)
c_2	780 (800)	220 (200)

Il valore χ^2 è calcolato come segue:

$\chi^2 = \frac{(820-800)^2}{800} + \frac{(780-800)^2}{800} + \frac{(180-200)^2}{200} + \frac{(220-200)^2}{200} =$
$\frac{1}{2} + \frac{1}{2} + 2 + 2 = 5.$

Quindi la probabilità $P(\chi^2_{0.05,1} \geq \chi^2) < 0.05$, poiché il valore χ^2 supera il valore critico $3,84$. Quindi la differenza di prestazioni è statisticamente significativa ($p < 0.05$).

Esercizio 3.1.

a) Qualsiasi maschera matriciale M tale che:

$$m_{ij} = \begin{cases} 1 & \text{se } (i_1, j_1) \leq (i, j) \leq (i_2, j_2), \\ 0 & \text{altrimenti.} \end{cases}$$

purché $(i_1, j_1) \leq (i_2, j_2)$ e $\leq$ è un ordinamento lessicografico. Ad esempio, una possibile maschera è la seguente:

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{pmatrix}$$

dove $(i_1, j_1) = (2, 1)$ e $(i_2, j_2) = (3, 1)$.

b) Dato il grafo $G = (V, E)$, $Fitness(h) = NumeroArchi(k, |E|) \cdot MaxConn(G)$, dove $NumeroArchi(k, |E|) = \frac{1}{|k-|E||+1}$ e $MaxConn(G) = \max_k |V_k|$, dove V_k sono le componenti connesse del grafo G. Altre soluzioni sono possibili, ad esempio: $MaxConn(G) = |\{(u, v) \in V \times V : \text{esiste un cammino tra u e v}\}|$.

c) i) Si sceglie $(1-r)p = \frac{1}{3} \cdot 3 = 1$ individuo da inserire in P_S (ricordiamo che $p = |P|$). Ne scegliamo uno:

$$P_S = \{ \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \}$$

ii) Si selezionano $\frac{rp}{2} = \frac{\frac{2}{3} \cdot 3}{2} = 1$ coppie da incrociare geneticamente: selezioniamo i primi due individui della popolazione originaria P e applichiamo la maschera di crossover specificata al punto (a):

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & \underline{1} & \underline{1} & \underline{0} \\ \underline{0} & \underline{0} & 0 & 0 \end{pmatrix}, \begin{pmatrix} \underline{0} & \underline{0} & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{0} & \underline{0} \\ \underline{0} & 0 & 0 & 1 \\ 1 & 0 & \underline{1} & \underline{0} \end{pmatrix}$$

ottenendo i due offspring:

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

che vengono aggiunti a P_S:

$$P_S = \{ \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \}$$

iii) Infine si applica la mutazione di un singolo bit a 2/3 degli individui di P_S:

$$P_S = \{ \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & \underline{1} & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & \underline{0} & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \}$$

Notiamo come, al termine di un'iterazione dell'algoritmo genetico, $|P_S| = |P|$. Infatti, gli elementi che aggiungiamo a P_S sono $(1-r) \cdot p$ durante il passo (i) e $r \cdot p$ durante il passo (ii), ovvero in totale $((1-r)+r) \cdot p = p$.

Esercizio 4.1. I valori calcolati durante la prima iterazione sono i seguenti:
$Q(A1, \rightarrow) = 0 + \gamma \cdot 0 = 0$
$Q(A2, \downarrow) = -10 + \gamma \cdot 0 = -10$
$Q(B2, \downarrow) = 0 + \gamma \cdot 0 = 0$
$Q(C2, \rightarrow) = 0 + \gamma \cdot 0 = 0$
$Q(C3, \rightarrow) = 10 + \gamma \cdot 0 = 10$
$Q(C4, \rightarrow) = 0$

Procedendo a ritroso:
$Q(C3, \rightarrow) = 10 + \gamma \cdot 0 = 10$
$Q(C2, \rightarrow) = 0 + \gamma \cdot 10 = 9$
$Q(B2, \downarrow) = 0 + \gamma \cdot 9 = 8,1$
$Q(A2, \downarrow) = -10 + \gamma \cdot 8,1 = -10 + 7,29 = -2,71$
$Q(A1, \rightarrow) = 0 + \gamma \cdot 0 = 0$

Esercizio 5.1.
$P(sì)P(stato = buono|sì)P(incroci = molti|sì)P(limite = 90|sì) = \frac{4}{7} \cdot \frac{1}{4} \cdot \frac{3}{4} \cdot \frac{2}{4} = \frac{1}{56}$
$P(no)P(stato = buono|no)P(incroci = molti|no)P(limite = 90|no) =$

$\frac{3}{7} \cdot \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{63}$

per cui: $c_{max} = \underset{c \in \{no, sì\}}{\operatorname{argmax}} P(c) \prod_{i=1}^{3} P(a_i|c) = sì$

Quesito 5.2. Tali approcci forniscono predizioni probabilistiche, ovvero predizioni effettuate con un grado di confidenza tra 0 e 1. Ogni esempio di addestramento progressivamente decrementa o incrementa la probabilità stimata che una certa ipotesi sia corretta. Inoltre, nuove istanze possono essere classificare combinando predizioni di ipotesi multiple, ciascuna pesata con il suo grado di probabilità.