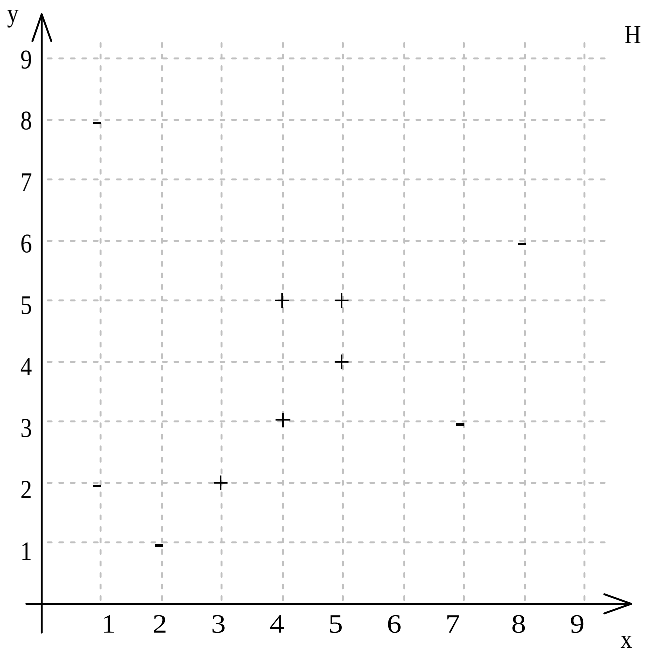
Apprendimento Automatico
Prova Intermedia A.A. 2007/2008
Prof. Roberto Navigli

1 Apprendimento di Concetti

Esercizio 1.1 (5 punti). Si consideri l'insieme delle istanze (x, y) sul piano bidimensionale dei numeri naturali, ovvero sia $X \subseteq \mathcal{N} \times \mathcal{N}$, e H lo spazio delle ipotesi tale che $h \in H$ è definita come il rettangolo sullo spazio bidimensionale che include le istanze positive (ovvero $h = (x_1, x_2, y_1, y_2)$ classifica come positive le istanze $(x, y) \in X$ per cui $x_1 \leq x \leq x_2$ e $y_1 \leq y \leq y_2$).

Sia dato l'insieme di addestramento $D = \{((3, 2), 1),\ ((4, 3), 1),\ ((5, 4), 1),\ ((4, 5), 1),\ ((5, 5), 1),\ ((1, 2), 0),\ ((2, 1), 0),\ ((8, 6), 0),\ ((7, 3), 0),\ ((1, 8), 0)\}$ (riportato in figura). Enunciare l'ipotesi massimamente specifica e massimamente generale appartenente allo spazio delle versioni $VS_{H,D}$. *Suggerimento:* non è necessario applicare l'algoritmo VersionSpace.



Quesito 1.2 (3 punti). Illustrare brevemente le differenze tra gli algoritmi FIND-S e VersionSpace.

2 Valutazione delle Prestazioni

Esercizio 2.1 (5 punti). A un classificatore $c : X \rightarrow \{0, 1\}$ viene fornito in input il seguente test set costituito da 10 istanze. Disegnare la matrice di confusione e determinare le prestazioni del classificatore sulle 10 istanze in termini di accuratezza, precisione, recall e F1 measure.

	classe reale	$c(d_i)$
d_1	0	1
d_2	0	0
d_3	1	1
d_4	1	1
d_5	0	1
d_6	1	1
d_7	0	0
d_8	1	1
d_9	1	1
d_{10}	1	0

3 Alberi di Decisione

Esercizio 3.1 (8 punti). Sia $X = \{0, 1\} \times \{0, 1\} \times \{0, 1\}$ l'insieme delle istanze rappresentate mediante 3 attributi booleani. Dato il seguente insieme di addestramento D, apprendere un albero di decisione mediante l'algoritmo ID3.

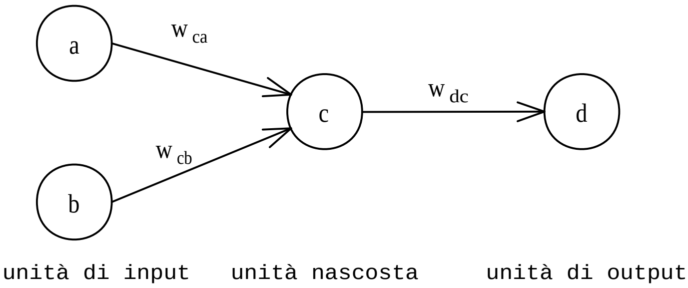
	A	B	C	$c(x)$
d_1	0	0	0	0
d_2	0	0	1	0
d_3	0	1	0	0
d_4	0	1	1	1
d_5	1	0	0	0
d_6	1	0	1	1
d_7	1	1	0	0
d_8	1	1	1	1

Quesito 3.2 (2 punti). In che senso possiamo dire che l'albero appreso è accurato?

Quesito 3.3 (3 punti). Come sarebbe l'albero di decisione se nell'insieme di addestramento non fosse incluso l'esempio d_2?
Suggerimento: non è necessario effettuare tutti i calcoli per rispondere a questa domanda.

4 Reti Neurali

Esercizio 4.1 (8 punti). Sia data la rete neurale alimentata in avanti in figura. Essa ha 5 pesi $(w_{c0}, w_{ca}, w_{cb}, w_{d0}, w_{dc})$. Supponiamo che i pesi siano inizializzati a $(0.1, 0.1, 0.1, 0.1, 0.1)$ e che il tasso di apprendimento $\eta = 0.3$.



Dato il seguente esempio di addestramento:

	a	b	c_d
d_1	1	0	1

Applicare i passi dell'algoritmo di BackPropagation per:

a. determinare il valore di o_c e o_d (l'output dell'unità nascosta c e dell'unità di output d);

b. determinare il valore di δ_d e δ_c;

c. aggiornare il peso di w_{ca} e w_{cb}.

Quesito 4.2 (3 punti). Come viene calcolato l'output di una singola unità percettrone a fronte di un esempio in input $\vec{x}$?

5 Soluzioni

Esercizio 1.1. $VS_{H,D} = \{(x_1, x_2, y_1, y_2) : 1 < x_1 \leq 3,\ 5 \geq x_2 < 7,\ y_1 = 2,\ y_2 = 5\}$. L'ipotesi massimamente specifica è $h_S = (3, 5, 2, 5)$, mentre esistevano almeno tre ipotesi massimamente generali (era sufficiente enunciarne una): $G = \{(2, 6, 2, +\infty), (3, 6, 0, +\infty), (2, +\infty, 2, 5)\}$ (accertabile anche applicando l'algoritmo VersionSpace).

Quesito 1.2. FIND-S trova l'ipotesi più specifica rispetto ai soli esempi positivi (eventualmente inconsistente con gli esempi negativi). VersionSpace, al contrario, identifica lo spazio delle versioni, ovvero il sottoinsieme dello spazio delle ipotesi che include tutte e sole le ipotesi consistenti con l'insieme d'addestramento D.

Esercizio 2.1. La matrice di confusione è la seguente:

	P	N
P	$TP = 5$	$FN = 1$
N	$FP = 2$	$TN = 2$

$$A = \frac{TP + TN}{TP + FP + TN + FN} = \frac{5 + 2}{5 + 2 + 2 + 1} = \frac{7}{10} = 0.70$$

$$P = \frac{TP}{TP + FP} = \frac{5}{5 + 2} = \frac{5}{7} = 0.714$$

$$R = \frac{TP}{TP + FN} = \frac{5}{5 + 1} = \frac{5}{6} = 0.833$$

$$F1 = \frac{2PR}{P + R} = \frac{2 \cdot 0.714 \cdot 0.833}{0.714 + 0.833} = \frac{1.1895}{1.547} = 0.768$$

Esercizio 3.1. Calcoliamo prima l'entropia della collezione D:

$$H(D) = -\frac{5}{8} log_2 \frac{5}{8} - \frac{3}{8} log_2 \frac{3}{8} = 0.954$$

Quindi, determiniamo l'entropia dei sottoinsiemi di D, uno per ogni scelta di valore per l'attributo A:

$$H(D_{A=0}) = -\frac{3}{4} log_2 \frac{3}{4} - \frac{1}{4} log_2 \frac{1}{4} = 0.811$$

$$H(D_{A=1}) = -\frac{2}{4} log_2 \frac{2}{4} - \frac{2}{4} log_2 \frac{2}{4} = 1$$

$$Gain(D, A) = H(D) - \sum_{v \in \{0,1\}} \frac{|D_{A=v}|}{|D|} H(D_{A=v}) = 0.954 - \frac{4}{8} \cdot 0.811 - \frac{4}{8} \cdot 1 = 0.954 - 0.4055 - 0.5 = 0.04$$

Per l'attributo B:

$$H(D_{B=0}) = -\frac{3}{4}log_2\frac{3}{4} - \frac{1}{4}log_2\frac{1}{4} = 0.811$$
$$H(D_{B=1}) = -\frac{2}{4}log_2\frac{2}{4} - \frac{2}{4}log_2\frac{2}{4} = 1$$
$$Gain(D, B) = 0.954 - \frac{4}{8} \cdot 0.811 - \frac{4}{8} \cdot 1 = 0.954 - 0.4055 - 0.5 = 0.04$$

Per l'attributo C:

$$H(D_{C=0}) = -\frac{4}{4}log_2 1 - \frac{0}{4}log_2 0 = 0$$
$$H(D_{C=1}) = -\frac{1}{4}log_2\frac{1}{4} - \frac{3}{4}log_2\frac{3}{4} = 0.811$$
$$Gain(D, C) = 0.954 - \frac{4}{8} \cdot 0 - \frac{4}{8} \cdot 0.811 = 0.954 - 0 - 0.405 = 0.549$$

Il guadagno informativo maggiore si ottiene scegliendo C come primo attributo. Per ogni $d \in D_{C=0}$, la classificazione è sempre 0, quindi abbiamo subito una foglia. Per $C = 1$, dobbiamo determinare quale attributo tra A e B discrimina meglio. Abbiamo:

$$H(D_{C=1,A=0}) = H(D_{C=1,B=0}) = -\frac{1}{2}log_2\frac{1}{2} - \frac{1}{2}log_2\frac{1}{2} = 1$$
$$H(D_{C=1,A=1}) = H(D_{C=1,B=1}) = -\frac{0}{2}log_2\frac{0}{2} - \frac{2}{2}log_2\frac{2}{2} = 0$$
$$Gain(D_{C=1}, A) = Gain(D_{C=1}, B)$$

Entrambi gli attributi forniscono lo stesso guadagno informativo. Scegliamo arbitrariamente A. Ogni esempio in $D_{C=1,A=1}$ è classificato 1. Per $D_{C=1,A=0}$, invece, dobbiamo inserire un ulteriore nodo di decisione, l'attributo B, sulla base del quale restituire classe 0 (quando B = 0) oppure 1 (quando B = 1).

Quesito 3.2. L'albero è accurato perché è stato appreso a partire dall'insieme di tutti i possibili esempi in X ed emette la classificazione corretta per ciascuno di essi.

Quesito 3.3. Guardando l'insieme d'addestramento, è evidente che l'attributo da scegliere è sempre C, con una differenza notevole: questa volta la scelta di C discrimina già perfettamente tra la classe 0 e la classe 1. Quindi l'albero di decisione risulta costituito da un nodo C e due figli, 0 e 1, che restituiscono la rispettiva classe 0 e 1.

Esercizio 4.1. Calcoliamo l'output delle unità c e d:

$o_c = \sigma(w_{c0} + o_a \cdot w_{ca} + o_b \cdot w_{cb}) = \sigma(0,1 + 1 \cdot 0,1 + 0) = \sigma(0,2) = \frac{1}{1+e^{-0,2}} = 0,54$

$o_d = \sigma(w_{d0} + o_c \cdot w_{dc}) = \sigma(0,1 + 0,54 \cdot 0,1 + 0) = \sigma(0,1 + 0,05) = \sigma(0,15) = 0,53$

Calcoliamo il valore di δ_d e δ_c:

$\delta_d = o_d(1 - o_d)(t_d - o_d) = 0,53(1 - 0,53)(1 - 0,53) = 0,11$

$\delta_c = o_c(1 - o_c)w_{dc}\delta_d = 0,54(1 - 0,54) \cdot 0,1 \cdot 0,11 = 0,0027$

Infine determiniamo il valore aggiornato dei pesi w_{ca} e w_{cb}:

$\Delta w_{ca} = \eta\delta_c x_{ca} = 0,3 \cdot 0,0027 \cdot 1 = 0,0008$

$w_{ca}^{new} = w_{ca} + \Delta w_{ca} = \eta\delta_c x_{cd} = 0,1 + 0,0008 = 0,1008$

$\Delta w_{cb} = \eta\delta_c x_{cb} = 0,3 \cdot 0,0027 \cdot 0 = 0$

$w_{cb}^{new} = w_{cb} + 0 = 0,1$

Quesito 4.2. $o(\vec{x}) = sign(\vec{w} \cdot \vec{x})$.