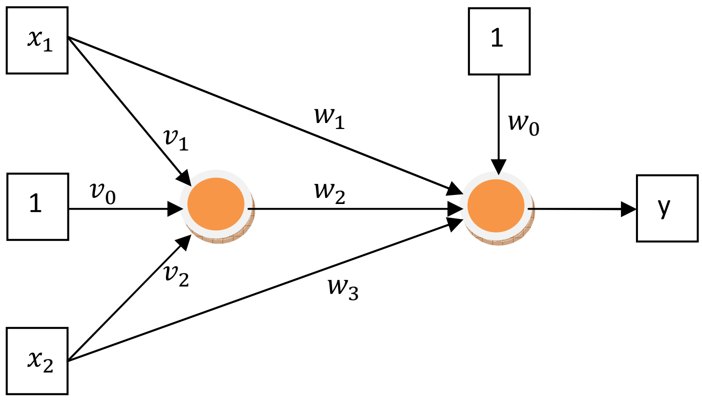
UNICAL - Facoltà di Ingegneria
Corso di Laurea Specialistica in Ingegneria Informatica
Corso di Data Mining e Scoperta di Conoscenza
Appello del 16/03/2011

Esercizio 1 (4 punti)

Si consideri la seguente rete neurale artificiale composta da due perceptron:



Si ricavi la formula di aggiornamento dei pesi della rete. In particolare, si definisca la formula per il peso v_1.

Esercizio 2 (6 punti)

Si consideri il seguente modello dato da una mistura di K distribuzioni:

$$p(\boldsymbol{x}) = \sum_{k=1}^{K} \pi_k p(\boldsymbol{x}|k)$$

Dove $p(\boldsymbol{x})$ è la probabilità della tupla $\boldsymbol{x}$, π_k è la probabilità a priori della k-esima distribuzione, mentre $p(\boldsymbol{x}|k)$ è la probabilità condizionata della tupla $\boldsymbol{x}$ data la k-esima distribuzione.

Si supponga che il vettore $\boldsymbol{x}$ possa essere suddiviso in due parti tali per cui $\boldsymbol{x} = \{\boldsymbol{x}_a, \boldsymbol{x}_b\}$. Dimostrare che la funzione della probabilità condizionata $p(\boldsymbol{x}_b|\boldsymbol{x}_a)$ è a sua volta una mistura di distribuzioni. Definire la complete data likelihood in termini di $p(\boldsymbol{x}_b|\boldsymbol{x}_a)$.

Esercizio 3 (3 punti)

Si consideri il seguente insieme di 15 item set frequenti di dimensione 3:

$$\{\{1\ 4\ 5\}, \{1\ 2\ 4\}, \{4\ 5\ 7\}, \{1\ 2\ 5\}, \{4\ 5\ 8\}, \{1\ 5\ 9\}, \{1\ 3\ 6\}, \{2\ 3\ 4\}, \{5\ 6\ 7\}, \{3\ 4\ 5\}, \{3\ 5\ 6\}, \{3\ 5\ 7\}, \{6\ 8\ 9\}, \{3\ 6\ 7\}, \{3\ 6\ 8\}\}$$

Calcolare il relativo hash tree, sapendo che una foglia dell'albero può contenere al massimo 3 item set, e che la funzione hash è la seguente:

$$hash(x) = x \bmod 3$$

ovvero il resto della divisione intera tra *x* e 3.

Soluzione

Esercizio 1

La funzione net è la seguente:

$$y = net(\boldsymbol{x}) = w_o + w_1 x_1 + w_3 x_2 + w_2(v_0 + v_1 x_1 + v_2 x_2)$$

La funzione di aggiornamento del vettore dei pesi $\boldsymbol{w}$ è:

$$\boldsymbol{w}^{(t+1)} = \boldsymbol{w}^{(t)} + \Delta\boldsymbol{w}^{(t)}$$

Dove, usando la funzione sigmoidale:

$$\Delta\boldsymbol{w}^{(t)} = \eta\big(class(\boldsymbol{x}) - net(\boldsymbol{x})\big)\boldsymbol{x}$$

con η tasso di learning.

Per ottenere in un processo iterativo un valore ottimale per il calcolo dei pesi, è sufficiente minimizzare il valore atteso dell'errore di predizione, definito come:

$$E[\boldsymbol{w}] = \frac{1}{2}\sum_{x}\big(class(\boldsymbol{x}) - net(\boldsymbol{x})\big)^2$$

Questo procedimento può essere svolto iterativamente, con l'obiettivo:

$$E\left[\boldsymbol{w}^{(t+1)}\right] \le E\left[\boldsymbol{w}^{(t)}\right]$$

Si noti che:

$$\Delta\boldsymbol{w}^{(t)} = \eta\nabla E\left[\boldsymbol{w}^{(t)}\right]$$

Tramite le serie di Taylor al primo ordine si ottiene:

$$E\left[\boldsymbol{w}^{(t+1)}\right] = E\left[\boldsymbol{w}^{(t)}\right] - \eta\left\|\nabla E\left[\boldsymbol{w}^{(t)}\right]\right\|^2$$

Volendoci concentrare su v_1, il calcolo del gradiente si riduce al calcolo di una derivata.

$$E\left[v_1^{(t+1)}\right] = E\left[v_1^{(t)}\right] - \eta\left(\frac{\partial}{\partial v_1}E\left[v_1^{(t)}\right]\right)^2$$

Dove

$$\frac{\partial}{\partial v_1}E\left[v_1^{(t)}\right] = \sum_{\boldsymbol{x}}\left[-x_1\big(class(\boldsymbol{x}) - net(\boldsymbol{x})\big)\right]$$

Esercizio 2

$$p(\boldsymbol{x}) = \sum_{k=1}^{K} \pi_k p(\boldsymbol{x}|k) = p(\boldsymbol{x}_b, \boldsymbol{x}_a) = \sum_{k=1}^{K} \pi_k p(\boldsymbol{x}_b, \boldsymbol{x}_a|k)$$

Ma:

$$p(\boldsymbol{x}_b, \boldsymbol{x}_a) = p(\boldsymbol{x}_b| \, \boldsymbol{x}_a)p(\, \boldsymbol{x}_a)$$

quindi:

$$p(\boldsymbol{x}_b| \, \boldsymbol{x}_a)p(\, \boldsymbol{x}_a) = \sum_{k=1}^{K} \pi_k p(\boldsymbol{x}_b| \, \boldsymbol{x}_a, k)p(\, \boldsymbol{x}_a|k)$$

Da cui:

$$p(\boldsymbol{x}_b| \, \boldsymbol{x}_a) = \frac{1}{p(\boldsymbol{x}_a)} \sum_{k=1}^{K} \pi_k p(\boldsymbol{x}_b| \, \boldsymbol{x}_a, k)p(\boldsymbol{x}_a|k) = \sum_{k=1}^{K} \frac{p(\boldsymbol{x}_b| \, \boldsymbol{x}_a, k)p(\boldsymbol{x}_a|k)\pi_k}{p(\boldsymbol{x}_a)}$$

Dal teorema di Bayes:

$$\frac{p(\boldsymbol{x}_a|k)\pi_k}{p(\boldsymbol{x}_a)} = p(k|\boldsymbol{x}_a)$$

Quindi:

$$p(\boldsymbol{x}_b| \, \boldsymbol{x}_a) = \sum_{k=1}^{K} p(\boldsymbol{x}_b| \, \boldsymbol{x}_a, k)p(k \mid \boldsymbol{x}_a)$$

Che è una mistura.

Procediamo con il calcolo della complete data likelihood.

$$L(\theta|\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) = \prod_{i=1}^{n} p(\boldsymbol{x}_i) = \prod_{i=1}^{n} \sum_{k=1}^{K} \pi_k p(\boldsymbol{x}_i|k)$$

Sfruttando le proprietà del logaritmo possiamo calcolare la log-likelihood

$$logL(\theta|\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) = \sum_{i=1}^{n} \log\left(\sum_{k=1}^{K} \pi_k p(\boldsymbol{x}_i|k)\right)$$

Introduciamo delle variabili latenti y_{ik} tali per cui $y_{ik} = 1$ se l'i-esima tupla è generata dalla k-esima componente, 0 altrimenti.

$$logL(\theta|\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) = \sum_{i=1}^{n} \log\left(\sum_{k=1}^{K} y_{ik} \pi_k p(\boldsymbol{x}_i|y_{ik}, k)\right)$$

Supponendo che ogni vettore $\boldsymbol{x}_i$ sia generato da una sola componente possiamo trasformare la log-likelihood in:

$$logL(\theta|\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) = \sum_{i=1}^{n}\sum_{k=1}^{K} y_{ik} \log\big(\pi_k p(\boldsymbol{x}_i|y_{ik}, k)\big) = \sum_{i=1}^{n}\sum_{k=1}^{K} y_{ik}\big(\log\big(p(\boldsymbol{x}_i|y_{ik}, k)\big) + \log(\pi_k)\big)$$

In termini di $p(\boldsymbol{x}_b|\boldsymbol{x}_a)$ si ottiene:

$$logL(\theta|\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) = \sum_{i=1}^{n}\sum_{k=1}^{K} y_{ik}\left(\log\left(p\left(\boldsymbol{x}_{b_i}, \boldsymbol{x}_{a_i}|y_{ik}, k\right)\right) + \log(\pi_k)\right)$$

$$= \sum_{i=1}^{n}\sum_{k=1}^{K} y_{ik}\left(\log\left(p\left(\boldsymbol{x}_{b_i}|\, \boldsymbol{x}_{a_i}, y_{ik}, k\right)\right) + \log\left(p\left(\boldsymbol{x}_{a_i}|y_{ik}, k\right)\right) + \log(\pi_k)\right)$$

UNICAL - Facoltà di Ingegneria
Corso di Laurea Specialistica in Ingegneria Informatica
Corso di Data Mining e Scoperta di Conoscenza
Appello del 16/03/2011

Esercizio 3