

mercoledì 7 dicembre 2022

14:26

1) Siano dati l'insieme delle categorie $C = \{\text{sport}, \text{politica}\}$ e una collezione di 280 documenti definiti sul vocabolario $V = \{T_1, T_2, T_3, T_4\}$.

a) Costruire un classificatore *bayesiano* per C , addestrandolo sul seguente training set TR:

TR = $\{ \langle D1, c2 \rangle, \langle D2, c2 \rangle, \langle D3, c1 \rangle, \langle D4, c1 \rangle \}$

dove per ogni documento si riporta di seguito l'elenco delle parole in esso presenti, con le relative occorrenze:

$D1 = \{T_4:4\}$ $D2 = \{T_3:3, T_2:2\}$

$D3 = \{T_2:2, T_4:4\}$ $D4 = \{T_3:5\}$

NB: illustrare chiaramente tutte le fasi di costruzione del classificatore

b) Determinare la classe di appartenenza del seguente documento $d = \{T_1:1, T_3:2\}$

(PUNTI 7)

(PUNTI 2)

Implementare un classif. KNN

		T_1	T_2	T_3	T_4
C_2 POLITICA	D1	0	0	0	4
	D2	0	2	3	0
C_1 SPORT	D3	0	2	0	4
	D4	0	0	5	0
?	D	1	0	2	0

Usare alg. Lechner

$$\vec{C}_{\text{POLIT}} = \frac{\vec{D}_1 + \vec{D}_2}{2}$$

$$\vec{C}_{\text{POLIT}} = \langle 0, 1, 1.5, 2 \rangle$$

$$\vec{C}_{\text{SPORT}} = \langle 0, 1, 2.5, 2 \rangle$$

$$\text{SIM}(\vec{D}, \vec{C}_{\text{POLIT}}) = 0 + 0 + 2 \times 1.5 + 2 \times 0 = 3$$

$$\text{SIM}(\vec{D}, \vec{C}_{\text{SPORT}}) = 5 \quad \leftarrow$$

Ranking

$$\begin{aligned} \text{SIM}(\vec{D}, \vec{D}_1) &= 0 \\ \text{SIM}(\vec{D}, \vec{D}_2) &= 6 \\ \text{SIM}(\vec{D}, \vec{D}_3) &= 0 \\ \text{SIM}(\vec{D}, \vec{D}_4) &= 10 \end{aligned}$$

SPORT D4
POLIT D2
POL D1
SP. D3

$$\vec{C}_{\text{SPORT}} = \frac{\vec{D}_3 + \vec{D}_4}{2}$$

mercoledì 7 dicembre 2022

14:50

- 1) Sia data la seguente matrice *utenti-item* di un sistema di filtraggio collaborativo, i cui rating di gradimento sono espressi in una scala discreta da 1 a 7

	I ₁	I ₂	I ₃	I ₄	I ₅
U ₁		3	5		7
U ₂		6	2	?	1
U ₃	3	4		1	2
U ₄		4	4	2	3

Calcolare la predizione per l'item I₄ e l'utente U₂, utilizzando un algoritmo di *user-to-user collaborative filtering*, una *neighborhood size* pari a 2 e il *coseno* come misura di similarità.

$$p(u_2, I_4) = \bar{r}_u + \frac{\sum_{j \in N(u)} \text{sim}(u_i, j) (r_{ji} - \bar{r}_j)}{\sum \text{sim}}$$

$$\text{sim}(u_2, u_1) = \text{Non Serve}$$

$$\text{sim}(u_2, u_3) = (6 \times 4 + 1 \times 2) / \|u_2\| \|u_3\| = 26 / \| \|$$

$$\text{sim}(u_2, u_4) = (6 \times 4 + 2 \times 4 + 1 \times 3) / \|u_2\| \|u_4\| = 35 / \| \|$$

$$\|u_2\| = \sqrt{36 + 4 + 1} = \sqrt{41} = 6.4$$

$$\|u_3\| = \sqrt{9 + 16 + 1 + 4} = \sqrt{30} = 5.4$$

$$\|u_4\| = \sqrt{16 + 16 + 4} = \sqrt{36} = 6.7$$

$$\text{sim}(u_2, u_3) = \frac{26}{6.4 \times 5.4} = 0.7$$

$$\text{sim}(u_2, u_4) = \frac{35}{6.4 \times 6.7} = 0.8$$

$$p(u_2, I_4) = 3 + \frac{0.7(1 - 10/4) + 0.8(2 - 13/4)}{0.7 + 0.8} =$$

$$= 3 + \frac{0.7(1 - 2.5) + 0.8(2 - 3.25)}{1.5} = 1.7$$