

Examen TD –ML–

(Documents et téléphones portables non autorisés)

Question : (3 points)

1. Comparer entre l'apprentissage supervisé et non supervisé
2. Pourquoi le classifieur Bayésien est-il «naïve»?

Exercice : (5 points)

Un expert médical veut construire un système de raisonnement basé sur des cas (des personnes) pour les tâches de diagnostic. Ces cas sont constitués d'un certain nombre de caractéristiques décrivant les symptômes possibles et leurs diagnostics (classification de la maladie). La base des cas contient les sept cas fournis dans le tableau ci-dessous.

Training	Fever	Vomiting	Diarrhea	Shivering	Classification
c ₁	no	no	no	no	healty (H)
c ₂	average	no	no	no	influenza (I)
c ₃	high	no	no	yes	influenza (I)
c ₄	high	yes	yes	no	salmonella poisoning (S)
c ₅	average	no	yes	no	salmonella poisoning (S)
c ₆	no	yes	yes	no	bowel inflammation (B)
c ₇	average	yes	yes	no	bowel inflammation (B)

En outre, l'expert a spécifié une mesure de similarités suite à son expertise, en utilisant les mesures locales de similarité et les poids des caractéristiques comme spécifié dans le tableau suivant.

sim _F				sim _V =sim _D =sim _{Sh}			Weights
q \ c	no	avg	high	q \ yes	no		
no	1.0	0.7	0.2	yes	1.0	0.0	w _F =0.3
avg	0.5	1.0	0.8	no	0.2	1.0	w _V =0.2
high	0.0	0.3	1.0				W _D =0.2
							w _{Sh} =0.3

Nous voulons classier le cas suivant q = (high; no; no; no) en utilisant le kNN avec un K=1.
 Calculer les similarités et donner sa classe.

Corrigé

Question :

Réponse: Naïve Bayes suppose que tous les attributs sont: 1) de même importance égal et 2) indépendants l'un de l'autre dans la classe.

Exercice:

$$Sim(q, c) = \sum_{a \in \{F, V, D, Sh\}} w_a \cdot sim_a(q_a, c.p_a)$$

(a) Calculate the similarity between all cases from the case base and the query $q = (high, no, no, no)$.

- ▶ for $c_1 = ((no, no, no, no), H)$:
 $Sim(q, c_1) = 0.3 \cdot 0.0 + 0.2 \cdot 1.0 + 0.2 \cdot 1.0 + 0.3 \cdot 1.0 = 0.70$
- ▶ for $c_2 = ((average, no, no, no), I)$:
 $Sim(q, c_2) = 0.3 \cdot 0.3 + 0.2 \cdot 1.0 + 0.2 \cdot 1.0 + 0.3 \cdot 1.0 = 0.79$
- ▶ for $c_3 = ((high, no, no, yes), I)$:
 $Sim(q, c_3) = 0.3 \cdot 1.0 + 0.2 \cdot 1.0 + 0.2 \cdot 1.0 + 0.3 \cdot 0.2 = 0.76$
- ▶ for $c_4 = ((high, yes, yes, no), S)$:
 $Sim(q, c_4) = 0.3 \cdot 1.0 + 0.2 \cdot 0.2 + 0.2 \cdot 0.2 + 0.3 \cdot 1.0 = 0.68$
- ▶ for $c_5 = ((average, no, yes, no), S)$:
 $Sim(q, c_5) = 0.3 \cdot 0.3 + 0.2 \cdot 1.0 + 0.2 \cdot 0.2 + 0.3 \cdot 1.0 = 0.63$
- ▶ for $c_6 = ((no, yes, yes, no), B)$:
 $Sim(q, c_6) = 0.3 \cdot 0.0 + 0.2 \cdot 0.2 + 0.2 \cdot 0.2 + 0.3 \cdot 1.0 = 0.28$
- ▶ for $c_7 = ((average, yes, yes, no), B)$:
 $Sim(q, c_7) = 0.3 \cdot 0.3 + 0.2 \cdot 0.2 + 0.2 \cdot 0.2 + 0.3 \cdot 1.0 = 0.47$