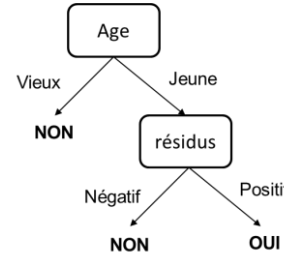


Corrigé Type

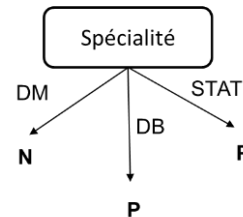
Exercice1 : Soient les informations des symptômes et du diagnostic des patients suivants. **Question :** Donner le modèle de décision en utilisant la méthode ID3. ⑥

Patient	Age	Résidus	Tension	Infection
01	Jeune	Positif	Élevée	oui
02	Jeune	Négatif	Élevée	non
03	Vieux	Positif	Élevée	non
04	Vieux	Positif	Normale	non
05	Jeune	Positif	Élevée	oui
06	Jeune	Négatif	Normale	non



Exercice2 : Le tableau suivant présente la base collectée pour l'entraînement. Questions : Construire un arbre de décision en utilisant le GiniIndex. ⑥

N°	Formation	Spécialité	Classement	Soutenance
1	LMD	DM	A	N
2	Ing	DM	B	N
3	LMD	STAT	A	P
4	Ing	STAT	B	P
5	LMD	DB	B	P
6	Ing	DM	A	N
7	LMD	DB	A	P



Etudiant caractérisé par : (Ing,DB,A). **Prediction impossible**

Questions Divers :

1) Définir c'est quoi un apprentissage automatique ?

C'est une discipline émergente pour développer une intelligence artificielle capable :

- d'analyser automatiquement des données ;
- de détecter des motifs et les utiliser des motifs pour la prédiction ;
- de prendre des décisions automatiques ;

2) Quatre problèmes qui peuvent être traités par des techniques d'apprentissage automatique.

Reconnaissance des formes, la vision artificielle, la classification de documents textuels, Détection de fraudes bancaires ... etc.

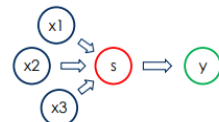
3) Type d'algorithme d'apprentissage pour segmenter vos clients en plusieurs groupes ? Clustering.

4) Le problème de détection du spam peut-il être classé comme un problème d'apprentissage supervisé ou non supervisé ? supervisé.

Exercice1 : RESEAUX DE NEURONES ②

Soit le neurone artificiel avec 3 entrées et 1 sortie :
Calculez la réponse y à l'excitation :

x = (5 2 1) et les poids w = (1 -5 2).



$$S = w_1x_1 + w_2x_2 + w_3x_3 = \sum w_i x_i = \begin{pmatrix} 5 \\ 2 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ -5 \\ 2 \end{pmatrix} = 5 \cdot 1 + 2 \cdot (-5) + 1 \cdot 2 = -3$$

Exercice 2 : Machines à vecteurs de support (SVM) :

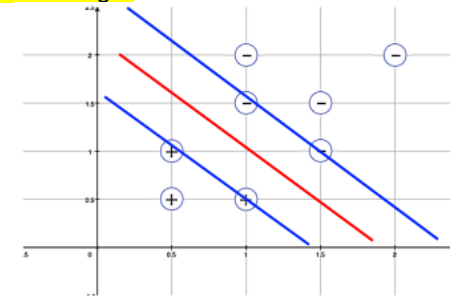
1. Quel est le but de l'algorithme SVM? ③

Les SVM sont des classificateurs linéaires qui cherchent un hyperplan pour séparer deux classes de données, positive et négative

2. Les frontières de décision sont linéaires pour la plupart des ensembles de données réelles. Comment traiter la cas de la non-linéarité par les SVM?

L'idée est de transformer les données d'entrée non linéairement séparables en un autre espace (généralement de plus grande dimension). Une frontière de décision linéaire peut séparer des exemples positifs et négatifs dans l'espace transformé.

3. Trouvez le SVM linéaire d'une façon graphique qui sépare de manière optimale les classes en maximisant la marge.



Exercice 3 : K-Means ③

Initialisation :

des centres de gravité : $\mu_1=8$ $\mu_2=10$ $\mu_3=11$

des clusters : $C_1=\emptyset$ $C_2=\emptyset$ $C_3=\emptyset$

Itération 1 :

Calcul des distances

Nombre 2 :

$d(2, \mu_1)=|2-8|=6$ $d(2, \mu_2)=|2-10|=8$ $d(2, \mu_3)=|2-11|=9$

2 est affecté au cluster C1.

Nombre 5 :

$d(5, \mu_1)=|5-8|=3$ $d(5, \mu_2)=|5-10|=5$ $d(5, \mu_3)=|5-11|=6$

5 est affecté au cluster C1.

Nombre 8 :

$d(8, \mu_1)=|8-8|=0$ $d(8, \mu_2)=|8-10|=2$ $d(8, \mu_3)=|8-11|=3$

8 est affecté au cluster C1.

Nombre 10 :

Corrigé Type

$d(10, \mu_1) = |10-8| = 2$ $d(10, \mu_2) = |10-10| = 0$ $d(10, \mu_3) = |10-11| = 1$

10 est affecté au cluster C2.

Nombre 11 :

$d(11, \mu_1) = |11-8| = 3$ $d(11, \mu_2) = |11-10| = 1$ $d(11, \mu_3) = |11-11| = 0$

11 est affecté au cluster C3.

Nombre 18 :

$d(18, \mu_1) = |18-8| = 10$ $d(18, \mu_2) = |18-10| = 8$ $d(18, \mu_3) = |18-11| = 7$

18 est affecté au cluster C3.

Nombre 20 :

$d(20, \mu_1) = |20-8| = 12$ $d(20, \mu_2) = |20-10| = 10$ $d(20, \mu_3) = |20-11| = 9$

20 est affecté au cluster C3.

Mise à jour des clusters :

$C1 = \{2, 5, 8\}$ $C2 = \{10\}$ $C3 = \{11, 18, 20\}$

R- estimation des centres de gravité :

$\mu_1 = (2+5+8)/3$ $\mu_2 = 10/1$ $\mu_3 = (11+18+20)/3$

$\mu_1 = 5$ $\mu_2 = 10$ $\mu_3 = 16.33$

Itération 2 :

Calcul des distances

Nombre 2 :

$d(2, \mu_1) = |2-5| = 3$ $d(2, \mu_2) = |2-10| = 8$ $d(2, \mu_3) = |2-16.33| = 14.33$

2 est affecté au cluster C1.

Nombre 5 :

$d(5, \mu_1) = |5-5| = 0$ $d(5, \mu_2) = |5-10| = 5$ $d(5, \mu_3) = |5-16.33| = 11.33$

5 est affecté au cluster C1.

Nombre 8 :

$d(8, \mu_1) = |8-5| = 3$ $d(8, \mu_2) = |8-10| = 2$ $d(8, \mu_3) = |8-16.33| = 8.33$

8 est affecté au cluster C2.

Nombre 10 :

$d(10, \mu_1) = |10-5| = 5$ $d(10, \mu_2) = |10-10| = 0$ $d(10, \mu_3) = |10-16.33| = 6.33$

10 est affecté au cluster C2.

Nombre 11 :

$d(11, \mu_1) = |11-5| = 6$ $d(11, \mu_2) = |11-10| = 1$ $d(11, \mu_3) = |11-16.33| = 5.33$

11 est affecté au cluster C2.

Nombre 18 :

$d(18, \mu_1) = |18-5| = 13$ $d(18, \mu_2) = |18-10| = 8$ $d(18, \mu_3) = |18-16.33| = 1.67$

18 est affecté au cluster C3.

Nombre 20 :

$d(20, \mu_1) = |20-5| = 15$ $d(20, \mu_2) = |20-10| = 10$ $d(20, \mu_3) = |20-16.33| = 3.67$

20 est affecté au cluster C3.

Mise à jour des clusters :

$C1 = \{2, 5\}$ $C2 = \{8, 10, 11\}$ $C3 = \{18, 20\}$

R- estimation des centres de gravité :

$\mu_1 = (2+5)/2$ $\mu_2 = (8+10+11)/3$ $\mu_3 = (18+20)/2$

$\mu_1 = 3.5$ $\mu_2 = 9.66$ $\mu_3 = 19$

Itération 3 :

Calcul des distances

Nombre 2 :

$d(2, \mu_1) = |2-3.5| = 1.5$ $d(2, \mu_2) = |2-9.66| = 7.66$ $d(2, \mu_3) = |2-19| = 17$

2 est affecté au cluster C1.

Nombre 5 :

$d(5, \mu_1) = |5-3.5| = 1.5$ $d(5, \mu_2) = |5-9.66| = 4.66$ $d(5, \mu_3) = |5-19| = 14$

5 est affecté au cluster C1.

Nombre 8 :

$d(8, \mu_1) = |8-3.5| = 4.5$ $d(8, \mu_2) = |8-9.66| = 1.66$ $d(8, \mu_3) = |8-19| = 11$

8 est affecté au cluster C2.

Nombre 10 :

$d(10, \mu_1) = |10-3.5| = 6.5$ $d(10, \mu_2) = |10-9.66| = 0.34$ $d(10, \mu_3) = |10-19| = 9$

10 est affecté au cluster C2.

Nombre 11 :

$d(11, \mu_1) = |11-3.5| = 7.5$ $d(11, \mu_2) = |11-9.66| = 1.34$ $d(11, \mu_3) = |11-19| = 8$

11 est affecté au cluster C2.

Nombre 18 :

$d(18, \mu_1) = |18-3.5| = 14.5$ $d(18, \mu_2) = |18-9.66| = 8.34$ $d(18, \mu_3) = |18-19| = 1$

18 est affecté au cluster C3.

Nombre 20 :

$d(20, \mu_1) = |20-3.5| = 16.5$ $d(20, \mu_2) = |20-9.66| = 10.34$ $d(20, \mu_3) = |20-19| = 1$

20 est affecté au cluster C3.

Mise à jour des clusters :

$C1 = \{2, 5\}$ $C2 = \{8, 10, 11\}$ $C3 = \{18, 20\}$

R- estimation des centres de gravité :

$\mu_1 = (2+5)/2$ $\mu_2 = (8+10+11)/3$ $\mu_3 = (18+20)/2$

$\mu_1 = 3.5$ $\mu_2 = 9.66$ $\mu_3 = 19$

Stabilité : Les centres de gravité n'ont pas changé. L'algorithme s'arrête

Exercice 4: KNN

On calcule la distance euclidienne qui sépare le point U à chacun des points :

Distance	Expression	Valeur
Distance(U, point 1)	$\text{Sqrt}((1-1)^2 + ((4-2)^2))$	02,00
Distance(U, point 2)	$\text{Sqrt}((2-1)^2 + ((6-2)^2))$	02,24
Distance(U, point 3)	$\text{Sqrt}((2-1)^2 + ((5-2)^2))$	01,41
Distance(U, point 4)	$\text{Sqrt}((2-1)^2 + ((1-2)^2))$	03,16
Distance(U, point 5)	$\text{Sqrt}((4-1)^2 + ((2-2)^2))$	03,61
Distance(U, point 6)	$\text{Sqrt}((5-1)^2 + ((6-2)^2))$	04,47
Distance(U, point 7)	$\text{Sqrt}((6-1)^2 + ((5-2)^2))$	05,10
Distance(U, point 8)	$\text{Sqrt}((6-1)^2 + ((4-2)^2))$	05,83

pour k (k=3) plus proches voisins sont les points (1, 2 et 3). Parmi ces 3 points, la classe C1 est majoritaire. **le point U sera affecté à la classe C1.**

Exercice 5 : Classifieurs de Bayes: ②

1. Pourquoi certains classifieurs de Bayes sont dits « naïfs » ?

parce qu'il suppose l'indépendance des attributs et par conséquent exprime la probabilité d'appartenance à une classe comme un produit de probabilités, mais en général ce n'est pas toujours le cas.

2. Quel est l'inconvénient majeur de la méthode de classification de Bayes.

suppose que les probabilités sont connues, mais dans la réalité ce n'est pas toujours le cas.