
Examen

Questions de cours (6 pts : 2 + 2 + 2)

1. Certaines références définissent l'apprentissage artificiel comme étant un jeu entre espaces. Donner les différents espaces utilisées en montrant la relation entre ces espaces.
2. Expliquer les notions de *risque empirique* et de *risque réel* en en précisant la différence.
3. Les méthodes d'apprentissage supervisé se divisent en deux types : classification et régression. Expliquer la différence entre les deux en donnant un exemple (une méthode) pour chaque type.

Exercices (14 pts : 4 + 4 + 4 + 2)

Notre département d'informatique désire utiliser l'apprentissage automatique afin d'améliorer le processus de sélection des étudiants pour l'accès à la première année master IOD pour la prochaine année 2016/2017. Pour atteindre cet objectif, le responsable de la spécialité IOD a décidé d'utiliser les résultats des étudiants inscrits en M1 IOD pour l'année en cours en se basant sur leurs résultats dans les matières : Algorithmique et Bases de données obtenus en Licence. La table suivante résume les données rassemblées pour l'entraînement :

N°	Algo (L2)	BDD (L2)	Admis M2 (2016/2017)
1	Excellent	Moyen	Oui
2	Faible	Excellent	Non
3	Moyen	Moyen	Non
4	Moyen	Excellent	Oui
5	Faible	Faible	Non
6	Excellent	Faible	Non

Exercice 1 Apprentissage par analogie (4 pts)

1. Calculer la décision estimée pour l'étudiant ayant les mentions (Algo : Moyen, BDD : Faible) par la méthode KPPV avec $K=5$ en utilisant la distance de Hamming généralisée :

$$\begin{cases} D(x_i, x_j) = 1 - \frac{1}{Nb_{Att}} \sum_{k=1}^{Nb_{Att}} \frac{f(x_{ik}, x_{jk})}{Nb_{modalités}} \\ avec f(x_1, x_2) = 1 \quad si \quad x_{ik} = x_{jk}, \quad 0 \quad sinon \end{cases}$$

Exercice 2 Apprentissage Bayésien (4 pts : 2 + 2)

1. Construire un modèle de décision bayésien en utilisant l'estimateur de Laplace.
2. Calculer la décision obtenus par ce modèle pour l'étudiant de l'exercice précédent.

Exercice 3 Arbre de décision (4 pts : 3 + 1)

1. Construire un arbre de décision à partir de la table d'entraînement en utilisant le gain d'information. On donne le tableau suivant représentant les valeurs de la fonction :

$$H(x, y) = -\frac{x}{x+y} \log_2\left(\frac{x}{x+y}\right) - \frac{y}{x+y} \log_2\left(\frac{y}{x+y}\right)$$

$x \backslash y$	1	2	3	4	5	6
6	0,592	0,811	0,918	0,971	0,994	1
5	0,650	0,863	0,954	0,991	1	
4	0,722	0,918	0,985	1		
3	0,811	0,971	1			
2	0,918	1				
1	1					

2. Trouver la décision pour l'étudiant de l'exercice 1 en utilisant l'arbre construit.

Exercice 4 Combinaison de modèles (2 pts)

- Trouver la décision pour l'étudiant de l'exercice 1 en utilisant la méthode de combinaison de modèle dite *Bagging* utilisant les trois modèles construits dans les exercices 1, 2 et 3.

★★★ Bonne chance ★★★

Dr A.Djeffal

Corrigé type

Questions de cours (6 pts)

1. Espaces utilisés :

- Espace de toutes les hypothèses possibles
- Espace des hypothèses considérées
- Espace de tous les exemples possibles
- Espace des exemples disponibles

L'apprentissage automatique consiste à rechercher dans l'espace des hypothèses considérées, l'hypothèse optimale sur les exemples disponibles.

2. Le risque réel représente l'écart entre l'hypothèse optimale dans l'espace de toutes les solutions possibles sur tous les exemples possibles et l'hypothèse trouvée.

Le risque empirique représente l'écart entre l'hypothèse trouvée et l'hypothèse optimale considérée sur les exemples disponibles.

Le risque réel est théorique alors que le risque empirique est utilisé pour l'apprentissage automatique

3. Les méthodes de classification sont utilisées pour prédire une classe discrète (symbolique ou nominale), par exemple, les arbres de décision.

les méthodes de régression sont utilisées pour prédire une classe continue (numérique), par exemple, la régression linéaire simple.

Exercices (14 pts : 4 + 4 + 4 + 2)

Exercice 1 : Apprentissage par analogie (4 pts)

1. $X = (\text{Moyen}, \text{Faible})$

$$D(X, X_1) = 1 - \frac{1}{2} \left(\frac{0}{3} + \frac{0}{3} \right) = \frac{6}{6}$$

$$D(X, X_2) = 1 - \frac{1}{2} \left(\frac{0}{3} + \frac{0}{3} \right) = \frac{6}{6}$$

$$D(X, X_3) = 1 - \frac{1}{2} \left(\frac{1}{3} + \frac{0}{3} \right) = \frac{5}{6}$$

$$D(X, X_4) = 1 - \frac{1}{2} \left(\frac{1}{3} + \frac{0}{3} \right) = \frac{5}{6}$$

$$D(X, X_5) = 1 - \frac{1}{2} \left(\frac{0}{3} + \frac{1}{3} \right) = \frac{5}{6}$$

$$D(X, X_6) = 1 - \frac{1}{2} \left(\frac{0}{3} + \frac{1}{3} \right) = \frac{5}{6}$$

Tri : $X_3, X_4, X_5, X_6, X_1, X_2$

\Rightarrow KPPV = (X_3 : non, X_4 : oui, X_5 : non, X_6 : non, X_1 : oui)

\Rightarrow Max = non

\Rightarrow **Décision = non**

Exercice 2 : Apprentissage Bayésien (4 pts : 2 + 2)

1. Modèle bayésien :

Classe	oui	non
	$\frac{2}{6}$	$\frac{4}{6}$
Algo		
Excellent	$\frac{2}{2}$	$\frac{2}{5}$
Moyen	$\frac{1}{5}$	$\frac{2}{7}$
Faible	$\frac{3}{5}$	$\frac{3}{7}$
BDD		
Excellent	$\frac{2}{2}$	$\frac{2}{2}$
Moyen	$\frac{1}{2}$	$\frac{1}{2}$
Faible	$\frac{3}{5}$	$\frac{3}{7}$

2. $X = (\text{Moyen}, \text{Faible})$

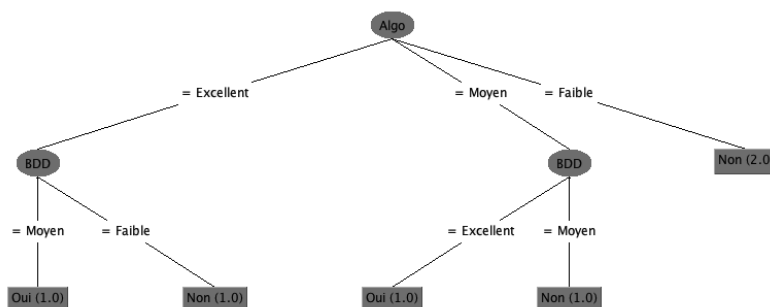
$$P(\text{oui}/X) = P(\text{oui}) \cdot P(X/\text{Oui}) = P(\text{oui}) \cdot P(\text{Moyen}/\text{oui}) \cdot P(\text{Faible}/\text{oui}) = \frac{2}{6} \times \frac{2}{5} \times \frac{1}{5} = \frac{4}{150} = 0.026$$

$$P(\text{non}/X) = P(\text{non}) \cdot P(X/\text{non}) = P(\text{non}) \cdot P(\text{Moyen}/\text{non}) \cdot P(\text{Faible}/\text{non}) = \frac{4}{6} \times \frac{2}{7} \times \frac{3}{7} = \frac{24}{294} = 0.081$$

\Rightarrow **Décision = non**

Exercice 3 : Arbre de décision (4 pts : 3 + 1)

1. Arbre :



2. $X = (\text{Moyen}, \text{Faible})$

\Rightarrow **Décision = ?**

Exercice 4 : Combinaison de modèles (2 pts)

– Bagging : Max vote \Rightarrow

Modèle 1 (KPPV) : non, Modèle 2 (Bayésien) : non, Modèle 3 : Arbre de décision : ?

\Rightarrow **Décision = non**