

Examen Réparti 1

Durée : 2H

Seuls documents autorisés : Supports et notes de cours

- Barème indicatif -

Exercice 1 (3pts) – Spectral clustering and graph Laplacian

You are given the adjacency matrix **A** and the degree matrix **D** :

$$\begin{array}{c} \mathbf{A} \\ \left[\begin{array}{cccc} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{array} \right] \end{array} \quad \begin{array}{c} \mathbf{D} \\ \left[\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{array} \right] \end{array}$$

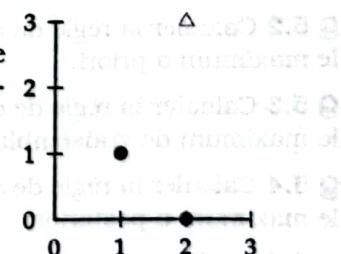
Q 1.1 Reconstruct the data graph based on **A** and **D** provided.

Q 1.2 What is the adjacency matrix? What is the degree matrix?

Q 1.3 Compute the Laplacian graph. How the Laplacian graph is used in the spectral clustering?

Exercice 2 (3pts) – Support vector machines

Consider building an SVM over a tiny data set of three points, where you have two classes (class \oplus : black circles; class \ominus : white triangles).



Q 2.1 Construct the decision boundary

Q 2.2 Estimate the margin between the classes

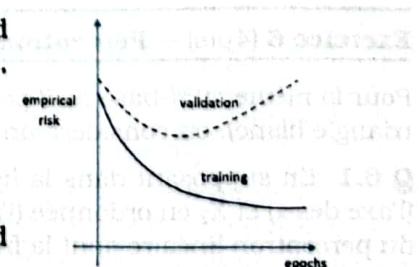
Exercice 3 (2pts) – Linear classification/regression. Training versus testing.

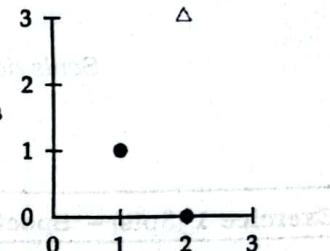
Q 3.1 Discuss the differences between linear regression and logistic regression (compare : the models, objective functions, optimisation).

Q 3.2 We trained a classifier and we got the following curves :

Q 3.2.1 Why the curves are different?

Q 3.2.2 What conclusion (for the training procedure) should we do?



Exercice 4 (3pts) - k-plus proches voisins

Pour la même mini-base de 3 points, proposer tous les classificateurs k-plus proches voisins possibles et dessiner leurs frontières.

Exercice 5 (7pts) - Classificateurs probabilistes, Naive Bayes

Soit un problème de classification binaire à 2 dimensions X_1 et X_2 (binaires elles aussi). Dans la base d'apprentissage, les représentants de la classe \ominus et \oplus se répartissent ainsi :

$C = \ominus$	X_1
X_2	0 1
0	30 30
1	15 25

$C = \oplus$	X_1
X_2	0 1
0	15 30
1	25 30

Par ailleurs, on considère que dans la population totale, la probabilité d'être dans la classe \ominus est de 60%, le reste de la population est donc dans la classe \oplus .

Q 5.1 Calculer les vraisemblances $P(X_1, X_2 | C = \ominus)$ et $P(X_1, X_2 | C = \oplus)$. Donner la loi a priori de la classe : $P(C = \oplus)$ et $P(C = \ominus)$.

Q 5.2 Calculer la règle de décision pour chaque valeur de (X_1, X_2) si la méthode utilisée est le maximum a priori.

Q 5.3 Calculer la règle de décision pour chaque valeur de (X_1, X_2) si la méthode utilisée est le maximum de vraisemblance.

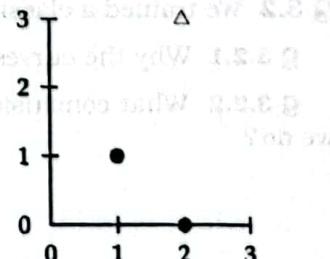
Q 5.4 Calculer la règle de décision pour chaque valeur de (X_1, X_2) si la méthode utilisée est le maximum a posteriori.

Q 5.5 Quelles sont les hypothèses d'un *Naive Bayes*? Calculer les vraisemblance avec cette hypothèse. Donner la règle de décision pour chaque valeur de (X_1, X_2) si la méthode utilisée est le maximum de vraisemblance d'un *Naive Bayes*.

Exercice 6 (4pts) - Perceptron

Pour la même mini-base de 3 points (classe \oplus : ronds noirs ; classe \ominus : triangle blanc), on considère un neurone formel à 2 entrées (x_1, x_2) ,

Q 6.1 En supposant dans la figure ci-contre que X_1 est en abscisse (l'axe des x) et X_2 en ordonnée (l'axe des y), donner la valeur des poids du perceptron linéaire dont la frontière est la droite d'équation $y = 0.5$ et qui a une erreur d'apprentissage de $\frac{1}{3}$.



Q 6.2 On prend $(w_0, w_1, w_2) = (1, 0, -2)$. Ces poids sont-ils compatibles avec votre réponse à la question précédente? Calculer une itération de l'algorithme d'apprentissage du perceptron linéaire en prenant epsilon = 1. Que pensez-vous de la classification obtenue?