
ÉVALUATION DU MODULE N° 5 : Classification

Consigne : Rédiger les réponses aux questions suivantes et rendre le travail sous forme d'un document `ipython notebook (.ipynb)`. Les résultats doivent être reproductibles, c'est-à-dire : le code doit être exécutable par le correcteur. Le fichier est à envoyer pour le 30 avril 2016, 23h59 dernier délai à l'adresse suivante : joseph.salmon@telecom-paristech.fr.

- COMPARAISONS DE DIFFÉRENTES MÉTHODES DE CLASSIFICATION -

On va comparer dans cette partie les différentes méthodes sur la base de donnée obtenue comme dans le premier TP avec les commandes suivantes :

```
from sklearn.datasets import load_digits
digits = load_digits()
X, y = digits.data, digits.target
```

On suivra le protocole expérimental suivant : couper les données en deux parties 79% pour l'apprentissage et 21% pour la validation (donner la taille des deux blocs choisis). Sur la partie d'apprentissage on entraînera les méthodes suivantes :

- Naive Bayes
- LDA
- Régression logistique
- QDA
- KNN (en prenant comme nombre de voisins $k = 1$)
- KNN (en choisissant k par validation croisée (V-fold) avec $V = 6$)
- une autre méthode de votre choix

On validera leur performance en donnant :

- 1) la proportion d'erreurs de classification faite sur la partie des données gardée pour la validation
- 2) le score F1.
- 3) Pour les méthodes mentionnées, proposer une synthèse sous forme de tableau ou de graphique, avec les renseignements suivants :
 - temps de calcul en seconde pris par chaque méthode pour la partie apprentissage (pour l'entraînement sur les 79% des données)
 - temps de calcul en seconde pris par chaque méthode pour la partie validation (sur les 21% restants)
 - pourcentage d'erreurs de classification de chaque méthode
 - le score F1.
- 4) On affichera les matrices de confusion associées : celles de la meilleure et de la pire des méthodes obtenues (au sens du nombre d'erreurs commises) parmi celles étudiées. Commentez vos résultats.

5) Proposer un (cours) paragraphe synthétisant l'ensemble de vos expériences ci-dessus.

- CRÉATION DE GRAPHE (SOUS MATPLOLIB PAR EXEMPLE) -

6) Écrire une fonction qui affiche le graphe (tridimensionnel) d'une densité de gaussienne en dimension deux et qui prend en entrée un vecteur de moyenne $mu = [mu_1, mu_2]$ et une matrice de covariance symétrique Σ (Sigma). On veillera à donner un message d'erreur "Matrice Sigma non-symétrique" dans le cas où la matrice Σ en entrée n'est pas symétrique. Insérer les graphes dans votre document pour :

$$\Sigma = \begin{bmatrix} 1/5 & 1/10 \\ 1/10 & 1/5 \end{bmatrix}$$

et

$$\Sigma = \begin{bmatrix} 1/5 & -2/10 \\ -2/10 & 3/5 \end{bmatrix}$$

Par simplicité on pourra prendre $mu_1 = mu_2 = 0$.