

- Durée : 2h
- Documents autorisés : cours et TD machine
- **A rendre : une archive contenant le script Matlab avec toutes les fonctions nécessaires pour l'exécuter**

L'entreprise PostalService dispose d'images de lettres et chiffres manuscrits. Elle propose de réaliser une fonction de classification simple et rapide de deux chiffres ("3" et "8") qu'elle a choisis.

Les images $I_i \in \mathbb{R}^{28 \times 28}$ sont transformées en vecteurs $x_i \in \mathbb{R}^{784}$. Ils sont regroupés dans $\mathcal{D} = \{(x_i, y_i) \in \mathbb{R}^{784} \times \{3, 8\}\}_{i=1}^N$. Ces données sont stockées dans $X_a \in \mathbb{R}^{N \times 784}$ et $Y_a \in \mathbb{R}^N$ et disponibles dans le fichier `datatrain.mat`.

Pour avoir un modèle simple la démarche retenue par PostalService est la suivante :

- projeter les données dans un espace de dimension réduite $q < 784$ en utilisant l'ACP
 - apprendre un modèle SVM linéaire sur les données projetées
1. Ecrire un programme Matlab qui réalise les opérations suivantes :
 - calculer la matrice de projection P de l'ACP
 - projeter les données dans un espace de dimension $q < 784$ en utilisant P
 - apprendre un modèle SVM linéaire $f(z) = w^\top z + b$ avec z la projection du point x par la matrice P
 - choisir la valeur appropriée de q avec $2 \leq q \leq 100$
 2. Evaluer le modèle sur les données d'apprentissage et de test en affichant les matrices de confusion et les erreurs de classification respectives. Les données de test sont disponibles dans le fichier `datatest.mat` sous la forme $X_t \in \mathbb{R}^{N_t \times 784}$, $Y_t \in \mathbb{R}^{N_t}$ et chaque étiquette $y_j \in \{3, 8\}$.

Remarque

- Pensez à normaliser les données projetées avant de faire l'apprentissage du modèle SVM