

TP : Analyse factorielle Discriminante

1 Chargement des données et pré-traitement

- Télécharger le fichier “digits.mat”.
- Charger ce fichier sous matlab en utilisant la fonction *load*. Le fichier contient les matrices suivantes :

- \mathbf{x} et \mathbf{xt} : matrices de données contenant respectivement $n = 3000$ et $nt = 1500$ exemples d’images manuscrites. Chaque ligne de ces matrices correspond à une image stockée sous la forme d’un vecteur transposé.

Les images sont de la forme suivante :



- \mathbf{y} et \mathbf{yt} : étiquettes des images décrites dans les matrices précédentes. Ce sont des vecteurs qui contiennent la classe (1, 7, 8) de chaque image de \mathbf{x} et \mathbf{xt} .
- Utiliser la fonction *reshape* pour extraire quelques images de taille 28×28 pour chaque classe. Les visualiser avec la fonction *imagesc*.

2 Analyse factorielle discriminante binaire

- Créer un problème de classification binaire à partir des trois classes. Vous pourrez par exemple choisir de classifier la classe 8 contre 1 et 7. Stocker les étiquettes binaires $(-1, 1)$ dans les vecteurs \mathbf{yb} et \mathbf{ytb} .
- Estimer les moyennes de chaque classes et les visualiser sous la forme d’images. En déduire la matrice Σ_b du problème binaire.
- Estimer les matrices de covariance Σ_k (fonction *cov*) pour chaque classe. En déduire la matrice Σ_w intra classe du problème binaire.
- Estimer la direction \mathbf{u} . Quel est le problème numérique et comment le résoudre ?
- Projeter les exemples d’apprentissage \mathbf{x} et de test \mathbf{xt} sur la direction \mathbf{u} . Superposer deux histogrammes montrant la séparation des classes. Conclusions ?
- Trouver un seuil permettant de prédire la classe à partir de la projection. Calculer le taux de bonne reconnaissance sur les données d’apprentissage et de test.
- Comparer les différentes projections \mathbf{u}_{rsb} , $\mathbf{u}_{\text{fisher}}$ et $\mathbf{u}_{\text{deflex}}$. Quelles sont leurs TBR.
- Refaire les étapes précédentes pour la détection de 1 contre 7 et 8 et 7 contre 1 et 8. Quelles sont les différences de performance ? À quoi sont-elles dues ?

3 Analyse multiclasse

- On travaille maintenant sur le cas multiclasse avec 3 classes.
- Estimer les moyennes et matrices de covariance de chaque classes.
- Calculer la projection \mathbf{U} à partir de la matrice $\Sigma_w^{-1}\Sigma_b$ (fonction *eigs*). Une fois encore il existe des problème numériques, comment les gérer ?

-
- Projeter les données d'apprentissage et de test sur le sous espace de dimension 2 et tracer les exemples. les classes sont elles bien séparées.
 - Comparer cette projection à une projection de type ACP (vue au TP précédent). Laquelle est la plus adapté pour une tâche de classification ?
 - Proposer une méthode de décision de la classe à partir de la projection 2D. Calculer les TBR pour l'es données d'apprentissage et de test.
 - Quel est l'effet de la régularisation sur les performances de prédiction ?