

Analyse, classification et indexation des données

Classification MAP vs classification linéaire avec et sans ACP

Archive fournie : `CC.zip` contenant les fichiers `donnees.mat` et `CC.m`.

Les réponses aux questions seront rendues sous la forme d'un court rapport. Vous rendrez également le code matlab qui aura produit vos résultats. L'ensemble est à envoyer par mail à `pcoupe@labri.fr` avant la fin de la séance (aucun retard ne sera accepté). Il faudra envoyer le rapport au format pdf avec pour nom de fichier `votreprenom_votrenom.pdf`. Le code et le rapport sont à envoyer sous forme d'une archive `.zip` avec pour nom d'archive `votreprenom_votrenom.zip`.

Dans ce sujet, on utilise d'une part un classifieur bayésien basé sur le maximum a posteriori et d'autre part un classifieur linéaire. L'objectif est de mesurer et d'expliquer les différences de résultat entre les classifieurs avec ou sans réduction de dimension. Vous utiliserez 10% des données pour entraîner vos modèles et le reste pour les tests. Les erreurs calculées seront données en pourcentages. Nous demandons l'erreur moyenne globale balancée comme résultat final. Le fichier `donnees.mat` fournit deux classes `C1` et `C2` et le fichier `CC.m` permet de les afficher.

Exercice 1.

1. Tout d'abord, visualisez les données avec le script fourni. Ensuite, donnez la dimension des observations ainsi que le nombre d'échantillons de chaque classe. Enfin, estimez les paramètres de la distribution de chaque classe (moyenne, etc.).
2. Comparez les performances du MAP et du classifieur linéaire basé sur les moindres carrés (tous les coefficients de b fixés à 1) pour la classification de `C1` et `C2`. Utilisez la boucle d'évaluation développée en TD (les 2 classifieurs sont dans la même boucle afin d'utiliser les mêmes échantillons de teste et d'entraînement). On utilisera 100 itérations. Commentez et discutez les résultats obtenus. Fournissez également une figure des erreurs commises en fonction des itérations.
3. Comparez les performances du MAP et du classifieur linéaire basé sur les moindres carrés pour la classification de `C1` et `C2` mais cette fois sélectionnez manuellement les 2 dimensions les plus discriminantes en comparant les différentes possibilités. Comme précédemment, utilisez la boucle d'évaluation développée en mettant les 4 classifications dans la même boucle. Commentez et discutez les résultats obtenus par rapport aux résultats précédents. Fournissez également une figure des erreurs commises en fonction des itérations.
4. Comparez les performances du MAP et du classifieur linéaire basé sur les moindres carrés pour la classification de `C1` et `C2` mais après réduction en 2 dimensions en utilisant l'ACP. Comme précédemment, utilisez la boucle d'évaluation développée en TD avec 100 itérations. Commentez et discutez les résultats obtenus par rapport aux résultats précédents. Fournissez également une figure des erreurs commises en fonction des itérations.
5. Comparez les performances du MAP et du classifieur linéaire basé sur les moindres carrés pour la classification de `C1` et `C2` mais après réduction en 2 dimensions en utilisant l'ACP mais cette fois centrez et réduisez les données. Comme précédemment, utilisez la boucle d'évaluation développée en TD avec 100 itérations. Commentez et discutez les résultats obtenus par rapport aux résultats

précédents sans normalisation. Fournissez également une figure des erreurs commises en fonction des itérations.

Exercice 2.

Répétez les mêmes expériences mais en utilisant cette fois uniquement 5% des données de chaque classe. Comparez les résultats obtenus sur ce jeu de données plus petit avec ceux obtenus sur le jeu de données complet. Commentez et discutez les résultats obtenus par rapport aux résultats précédents. Fournissez également une figure des erreurs commises en fonction des itérations.