

TP4 – Classification bayésienne

Dans ce TP, nous souhaitons résoudre un problème de classification binaire en utilisant une méthode de classification bayésienne dans laquelle, chaque classe va être modélisée en utilisant la méthode « Kernel density Estimation »

I. Chargement et visualisation des données

Charger les données (TP4.npy) et visualiser les. Combien y a-t-il de points dans la base d'apprentissage ? Dans la base de test ? Quelle est la dimension des données ?

II. Estimation des densités de probabilité

Ecrire une fonction `ddp_parzen(x, X, Cov)` qui estime la densité de probabilité en un point x à partir des points de la base d'apprentissage X , en utilisant un noyau 2D gaussien. On utilisera une matrice de covariance diagonale avec un écart type σ identique pour les deux dimensions. On pourra utiliser la fonction `norm2()` développée au TP3.

Questions

Rappeler le principe de fonctionnement de l'estimation de densité de probabilité par noyau. Que représente le noyau ?

Afin de visualiser les densités de probabilité (ddp), estimer la ddp en tout point d'une grille allant de 1 à 20 et la stocker dans une matrice que l'on pourra visualiser sous forme d'image (attention à retourner l'image avant de la visualiser, `np.flipud`). Visualiser la ddp de chacune des deux classes et faire varier σ .

Questions

Comment varient les ddp en fonction de σ ? Etait-ce prévisible ? A priori et sans faire le test, quelle valeur de σ pensez-vous adéquate pour faire la classification ?

III. Classification bayésienne

Réaliser la classification des points de la base de test en n'omettant pas d'utiliser les probabilités a priori. Afficher le taux de reconnaissance.

Faire varier σ et optimiser le taux de reconnaissance.

Questions

Pour quelle valeur de σ obtient-on la meilleure classification ? Etait-ce prévisible ? Expliquer le résultat.