

TP k-plus proches

October 29, 2015

1. charger et lire le notebook “knn_regression.ipynb”
2. Reprendre les données “arbres remarquables”
3. Évaluez la qualité du k-plus proches voisins pour $k=1..30$. Quel est le meilleur k ? pour la suite du TP, appelons le k^*
4. normalisez les données et recommencez
5. essayez l’algorithme d’apprentissage de distance présent dans knn_regression.ipynb. Qu’obtenez-vous ?
6. Essayez l’algorithme suivant:
 - (a) pour $t = 1$ à 100:
 - i. tirez une matrice P de dimension $d \times d$ aléatoire (par exemple selon la loi normale)
 - ii. projetez les données sur P (cela se fait par une multiplication de matrices)
 - iii. calculez le score de cross-validation avec $k=k^*$
 - (b) affichez le meilleur score obtenu pendant les itérations