

Comparaison

December 10, 2018

1 Comparaison des 3 méthodes: k-NN, ANN et SVM

Réaliser par Mohamed ELFILALI et Nguyen Duc Hau

1.1 Contexte

Ce travail est réalisé dans le cadre du cours Apprentissage, où on va expérimentées trois méthodes réparties en trois TP. Le jeu de donnée utilisé est **MNIST original**.

1.2 Comparaison

A partir de l'expérimentation (cf. **TP1.pdf**, **TP2.pdf**, **TP3.pdf**), nous en déduisons le tableau ci-dessous.

Critère	k-NN	ANN	SVM
Vitesse de convergence vers l'optimum	relativement vite	très lente	moyen
Réglementation pour bias et variance	paramètre k	paramètre alpha	paramètre C
Complexité de développement	Très facile à concevoir un modèle manuel	Très compliqué, surtout l'algorithme de backpropagation	Compliqué et demande une compréhension de données et de compétence mathématique pour implémenter les fonctions kernels
Complexité en production	Facile à suivre et debugger	Le fonctionnement dans les couches intermédiaire est une boîte noire donc obscure pour analyser le résultat	Si la nature des données est connue, cette méthode devient facile à ajuster les paramètres

Critère	k-NN	ANN	SVM
Intérêt	Facile à comprendre et mettre en place, haute vitesse de calcul	Apdaté aux problèmes de natures différents, donc intéressant pour les problèmes complexes	Permet d'obtenir un modèle ajusté très fin au problème, robuste aux données bruitées et/ou à grande dimension
Inconvénient	Gourmande en espace de mémoire, prédiction lente, sensible aux features non pertinentes ou corrélés	Difficile à interpréter le processus d'apprentissage comme son fonctionnement reste un boîte noire	Difficile à traiter un jeu de données en grande nombre d'observation (car la matrice de Gram doit être stockée), difficile à choisir une fonction noyau sans avoir une expertise de métiers
Exemple d'application	Les problèmes convexes, données bruitées, pas besoin une précision élevée (système de recommandation, etc.)	Les problème très complexes pour modéliser, un grand jeu de données, bruitées possible, nécessité d'adaptation pour la prédiction soit traitée en temps réel, identification des features pertinents.	Les problèmes typiques pharmaceutiques: grande dimension alors que le nombre d'observation est peu, besoin d'un modèle sur mesure.

A noter que, l'intérêt et l'inconvénient de chaque méthode ont été détaillé dans les rapports spécifiques (cf. [TP1.pdf](#), [TP2.pdf](#), [TP3.pdf](#)). Ce tableau récapitule les grandes lignes de ces champs.

2 Annexe

Vous trouverez le code exécutable sur le lien github suivant : <https://github.com/Kihansi95/apprentissage/tree/master/superviser>