

Résultats de recherche sur la classification supervisée

M. Deker Sylvain¹

Mme. Courdy-Bahsoun Clémence²

¹ M1 IGAI Université Paul Sabatier

² M1 IGAI Université Paul Sabatier

sylvain.deker@univ-tlse3.fr

clemence.courdy-bahsoun@univ-tlse3.fr

Résumé

Cet article vise à comparer les avantages et les désavantages des classifications supervisées par bayésienne avec gaussienne et par la méthode des k-plus proches voisins, en s'appuyant sur les résultats obtenus sur un même jeu de données.

Mots Clef

Apprentissage, classification bayésienne, classification k-ppv.

Abstract

Keywords

Learning, k-NN classification,.

1 Introduction

Classification bayésienne avec gaussienne et k-ppv sont des classifications supervisée dont l'objectif est de répartir les données par classe avec un taux d'erreur minimisé. Ici pour des jeux de donnée identiques nous avons réalisés différents tests sur les deux méthodes de classification afin d'en définir les avantages et les inconvénients respectifs. A l'issue de ces tests des recherches sur l'évaluation des données par validation croisée a été mené.

2 Classification bayésienne avec Gaussienne

La classification bayésienne avec Gaussienne est une minimisation de la vraisemblance qui vise à prédire la classe d'appartenance la plus probable d'une donnée. Néanmoins avec des données réelles il est difficile d'obtenir des résultats satisfaisants car les données ne sont pas simples. Ainsi pour améliorer le modèle on tient compte de la probabilité a priori en appliquant la loi de Bayes.

Dans les tps précédant, nous avons déjà implémenté en python l'algorithme de prédiction permettant de classer les données fournies. Ici nous sommes reparti de ce code qui était basé sur une simplification de la vraisemblance en log-vraisemblance. C'est-à-dire que la vraisemblance qui est

une expression exponentielle a été simplifié par l'application du logarithme népérien afin d'obtenir un calcul un peu plus optimisé. De plus toutes les constantes ne dépendant pas des données ont été supprimé, n'apportant pas une précision indispensable dans la reconnaissance des données.

3 Classification k-NN

4 Evaluation par validation croisée

5 Résultats

6 Conclusion