k-Nearest-Neighbor Classification and Regression using SAS [*]

Liang Xie
Travelers Insurance

March 21, 2012

Abstract

K-Nearest Neighbor (KNN) classification and regression are two widely used analytic methods in predictive modeling and data mining fields. They provide a way to model highly nonlinear decision boundaries, and to fulfill many other analytical tasks such as missing value imputation, local smoothing, etc.

In this paper, we discuss ways in SAS to conduct KNN classification and KNN Regression. Specifically, PROC DISCRIM is used to build multi-class KNN classification and PROC KRIGE2D is used for KNN regression tasks. Technical details such as tuning parameter selection, etc are discussed. We also discuss tips and tricks in using these two procedures for KNN classification and regression.

Examples are presented to demonstrate full process flow in applying KNN classification and regression in real world business projects.

1 Introduction

---

[*]Draft version