Machine Learning for Cities
CUSP-GX-5006-002

Prof. Luis Gustavo Nonato

Assignment II

Release date: Feb. 21st, 2017 Due date: Mar. 7th, 2017

Naive Bayes and SVM: a comparative study

Motivation

Classification is major task in machine learning and many different classification techniques are available. Naive-Bayes and SVM are among the most important classification techniques, being widely used. This assignment aims to compare those two techniques in a real application scenario, making clear their effectiveness and computational demand.

Description

For this assignment students should:

1. Perform classification using both Naive-Bayes and SVM. In particular, SVM should be used with Euclidean and kernel functions.

2. Training and test data should be chosen properly in order to assess the effectiveness of each classification scheme.

3. The accuracy of each regression model should be assessed via cross-validation;

4. Comparisons between the three classification schemes should be accomplished. It is highly encouraged to preform a numerical study as to the impact of the training size in the quality of the classification.

5. Results should be presented as a report (template in NYU Classes) using a combination of tables and graphics.

Data Set

The data set "`manhattan-dof.csv`" (downloadable from NYU Classes) can be used for classification, using the column `BldClassif` as classes. If other data set is used in the experiments, a clear description of its attributes (and the chosen dependent variable) should be included in the assignment report.