Comparison of various Classification Techniques over Datasets of Different Feature Distributions

Concepts in Statistical Learning Theory
CS6464

March 9, 2017

1 Problem Statement

Given P, N-dimensional data points from two classes, compare the classification accuracies for different classification algorithms (*viz.* Bayes, K-Nearest Neighbor, Support Vector Machines etc.).

2 Input

- $P = 0.7K$, $7K$, $70K$, $0.7M$

- $N = 3$, 10, 50, 100

- Input contains data from four different distributions for each combination of P and N (specified as scatter modes).

3 Output

- Classification accuracy obtained using the assigned algorithms.

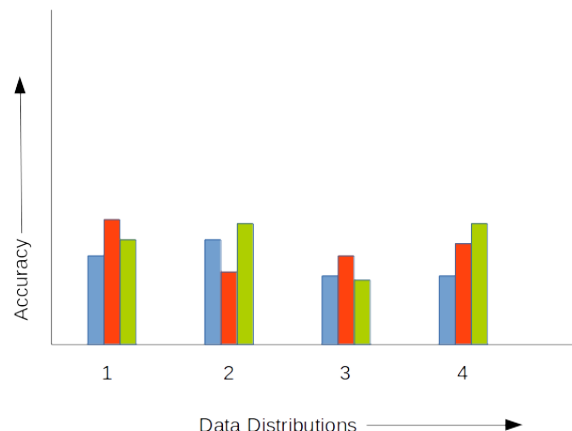- Plot (bar chart) of accuracies obtained by classifiers for each combination of P and N, for four different data distributions.

Figure 1: Example plot of the obtained accuracies.