

Comparison of various Classification Techniques over Datasets of Different Feature Distributions

Concepts in Statistical Learning Theory
CS6464

February 20, 2018

1 Problem Statement

Given N , P -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

- $N = 0.7K, 7K, 70K$
- $P = 3, 10, 50$

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.

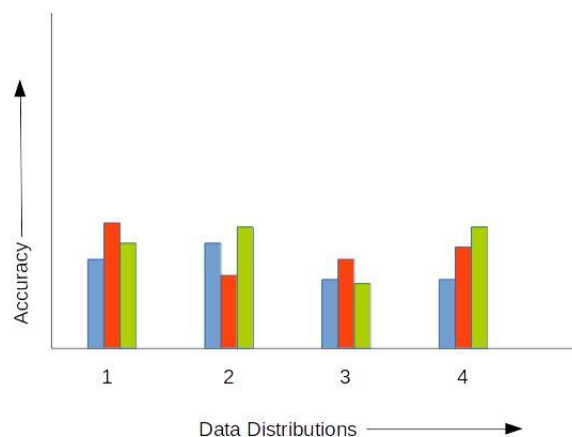


Figure 1: Example plot of the obtained accuracies.