## Data Mining

## Classification I

## 1 K-nearest neighbor classification

The objective is to study the use of the method k-nn and

- 1. Implement k-nn and test on the Iris dataset. Calculate the classification error? (by comparing the class labels obtained with the prediction and the original labels of test data)
- 2. Vary the value of k, comment on the results
- 3. Try to normalize the input dataset, is the performance better?
- 4. Apply PCA and SVD on the dataset, then what is the performance of k-nn on the new projected data? Justify the answer.
- 5. Propose an approach to improve the performance of k-nn with the help of k-cross validation.
- 6. Apply leave-one-out and calculate the error of classification.
- 7. Repeat the same process for two more datasets.

## 2 Perceptron classifier

In this problem, we aim to implement Perceptron classifier.

- Initialize values for the weight vector  $\mathbf{w}$  and fraction  $\alpha$
- Plot linear classifiers for Iris dataset (using PCA or SVD to reduce the number of dimensions to 2D)

- Comment on the convergence rate of Perceptron on Iris? How to make it converging faster?
- Try Perceptron classifier for 2 other datasets.