Statistical Learning
IIT Madras

Assignment 2

Submitted by:
Vivek Kumar Agrawal
Roll No.: CS17M094

1. Classifying data using KNN (K-Nearest Neighbours) and MKL (Multiple Kernel Learning) and comparing the accuracies obtained by both these classifiers :

The models are **trained** on **71.5%** data of each distribution of every DataSet and rest **28.5%** is used for **validation** as given below,

Size of DataSet (N)	**Training data size**	**Validation data size**
0.7k	**0.5k**	**0.2k**
7k	**5k**	**2k**
70k	**50k**	**20k**

1.1 Classification Accuracy obtained on Validation data using *KNN*:

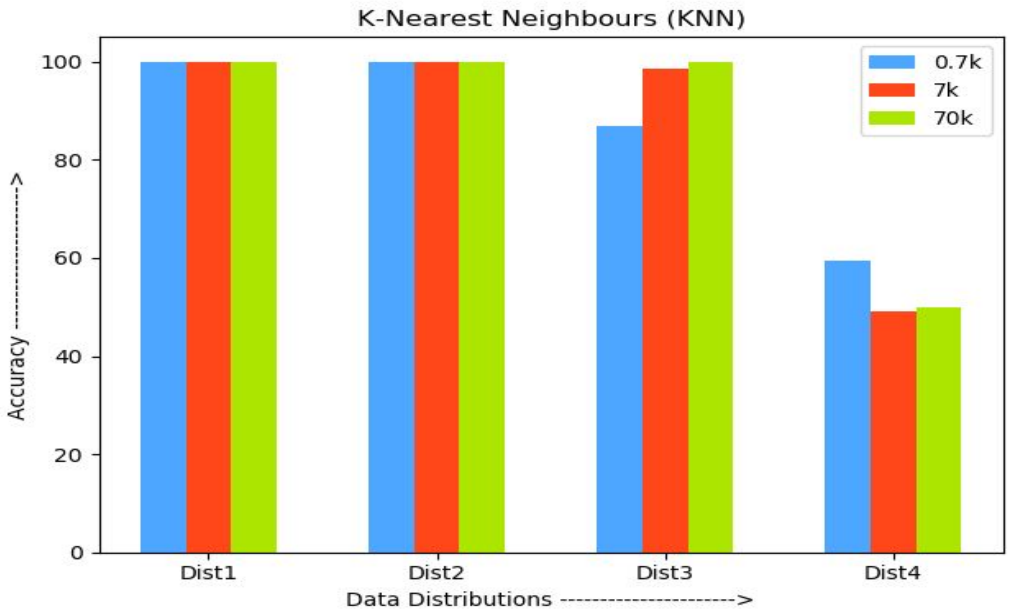
Size of DataSet (N)	**Dist1**	**Dist2**	**Dist3**	**Dist4**
0.7k	100	100	86.75	59.5
7k	100	100	98.1	49
70k	100	100	100	49.74

1.2 Classification Accuracy obtained on Validation data using *MKL*:

Size of DataSet (N)	**Dist1**	**Dist2**	**Dist3**	**Dist4**
0.7k	100	100	94.3	54.5
7k	100	100	99.6	51
70k	100	100	100	49

2. Bar Chart (Validation)

2.1 KNN



2.2 MKL

3. Accuracy obtained on Test Data

3.1 Classification Accuracy obtained on Test data using *KNN*:

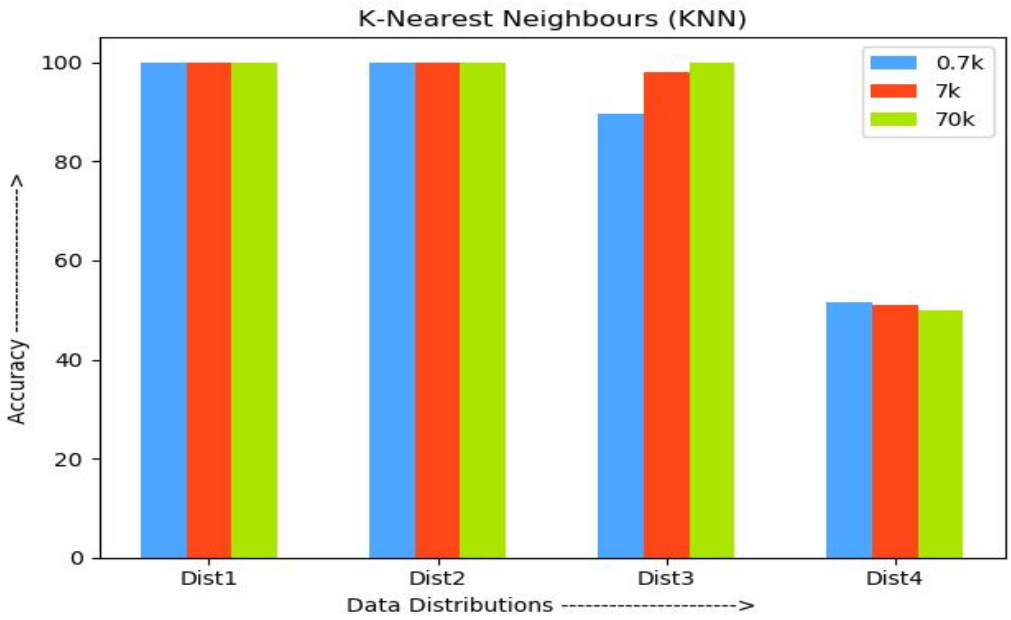
Size of DataSet (N)	**Dist1**	**Dist2**	**Dist3**	**Dist4**
0.7k	100	100	89.5	51.5
7k	100	100	98.05	51
70k	100	100	99.9	49.9

3.2 Classification Accuracy obtained on Test data using *MKL*:

Size of DataSet (N)	**Dist1**	**Dist2**	**Dist3**	**Dist4**
0.7k	100	100	94.15	45.6
7k	100	100	99.5	50.2
70k	100	100	100	49

4. Bar Chart (Test)

4.1 KNN



4.2 MKL