CS 5783 Assignment 1

Suraj Pawar

August 2019

1 kD Tree

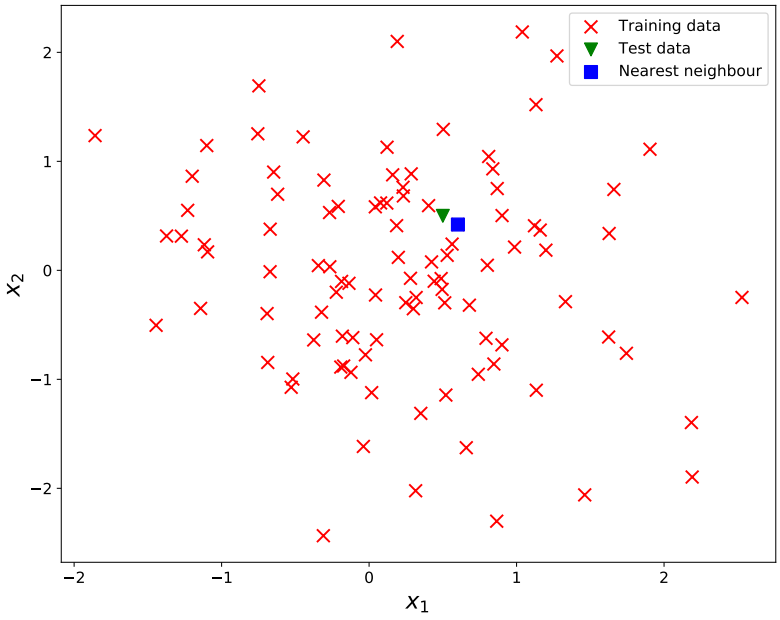

Figure 1: The training data consists of 100 points selected from random normal distribution. The test data is $[0.5, 0.5]$ and the nearest neighbour is correctly found.

2 Constructing a simple dataset

```
n1 = 5000
n2 = 5000
split = 0.5

mean1 = np.random.randn(1,2).flatten()
cov1 = make_spd_matrix(2,2)
data1 = np.random.multivariate_normal(mean1, cov1, n1)
label1 = np.zeros((n1,1))

mean2 = np.random.randn(1,2).flatten()
cov2 = make_spd_matrix(2,2)
data2 = np.random.multivariate_normal(mean2, cov2, n2)
label2 = np.ones((n2,1))

data = np.vstack((data1,data2))
labels = np.vstack((label1,label2))

# select indices randomly for splitting train and test data
indices = np.full(data.shape[0],True)
indices[:int(data.shape[0]*split)] = False
np.random.shuffle(indices)

# split the data into train and test data
xtrain, ytrain = data[indices==True], labels[indices==True]
xtest, ytest = data[indices==False], labels[indices==False]
```

3 Linear classifier

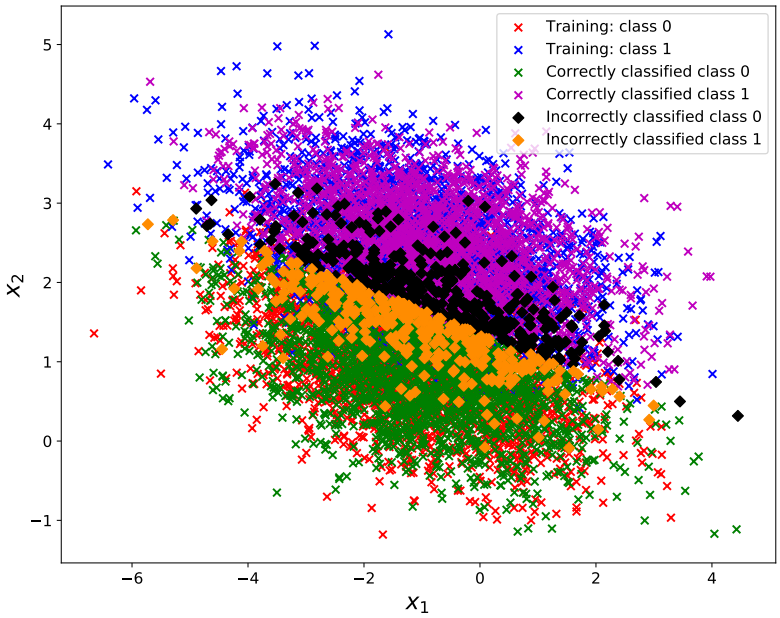

Figure 2: The data is generated from two different multivariate Gaussian distribution and is split into train and test data with 50%. Linear classifier achieves accuracy of **86.76%** on test data.

4 Nearest neighbour classification

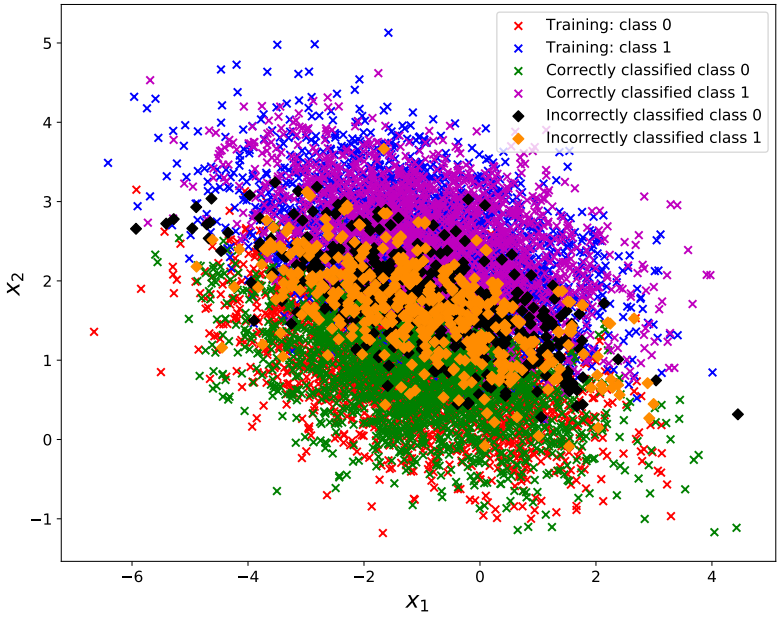

Figure 3: The data is generated from two different multivariate Gaussian distribution and is split into train and test data with 50%. kD tree classifier achieves accuracy of **82.32%** on test data.

5 Increasing complexity

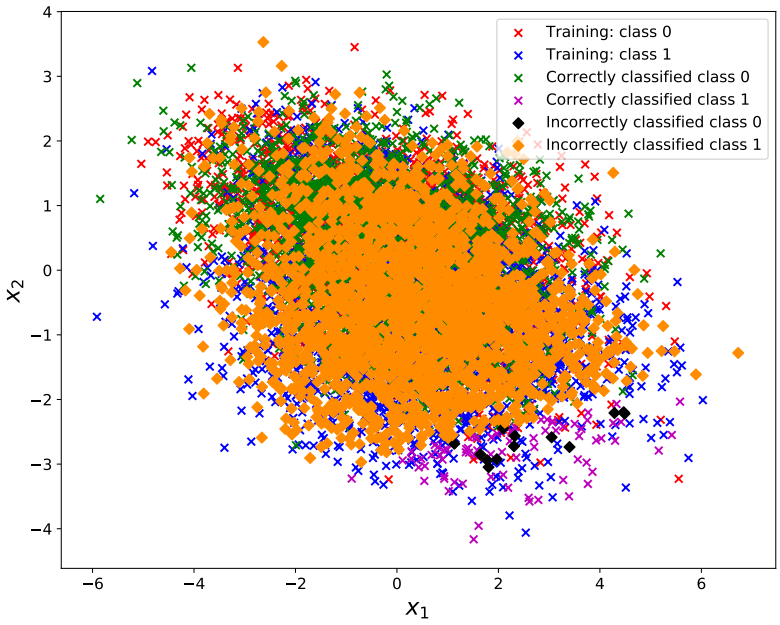

Figure 4: The data is generated from ten different multivariate Gaussian distribution and is split into train and test data with 50%. Five of them are assigned label 0 and the others 1. Linear classifier achieves accuracy of **51.2%** on test data.

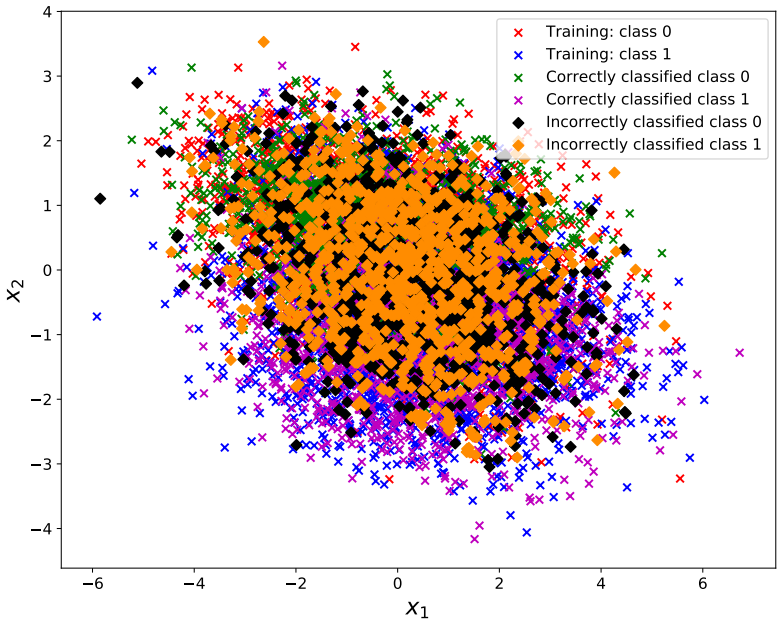

Figure 5: The data is generated from ten different multivariate Gaussian distribution and is split into train and test data with 50%. Five of them are assigned label 0 and the others 1. kD tree classifier achieves accuracy of **57.84%** on test data.