Report SMAI Assignment-1

Roll No :- 201525118

Question-1

For a 9000:3000 split in the training data:

- **Part - A**
-

  2997/3000 correctly classified
  Precision = 0.998
  Recall = 0.999
  Accuracy = 0.999

  Part - B

  For b=100000,
  2998/3000 correctly classified
  Precision = 0.999
  Recall = 0.999
  Accuracy = 0.999

  For b=100,
  2997/3000 correctly classified
  Precision = 0.998
  Recall = 0.999
  Accuracy = 0.999

  For b=5,
  2997/3000 correctly classified
  Precision = 0.998
  Recall = 0.999
  Accuracy = 0.999

  Part - C

  Without margin

  2998/3000 correctly classified
  Precision = 0.999
  Recall = 0.999
  Accuracy = 0.999

  With margin

For b=100000,
2998/3000 correctly classified
Precision = 0.999
Recall = 0.999
Accuracy = 0.999

Question-2

- **Part - A - Relaxation**
  Epochs - 5000 and b = 50
  Precision - 0.926829268293
  Accuracy - 0.97
  Recall - 1.0

  Epochs - 10000 and b = 5
  Precision - 0.926829268293
  Accuracy - 0.97
  Recall - 1.0

  Epochs - 10000 and b = 50
  Precision - 0.926829268293
  Accuracy - 0.97
  Recall - 1.0

  Epochs - 100 and b = 50
  Precision - 0.926829268293
  Accuracy - 0.97
  Recall - 1.0

- **Part - B - Updated Perceptron algorithm**

I updated the Algorithm such that n(eta) is changed to (1 - accuracy) for the iteration such that as number of errors increases, the value of n(eta) decreases.

Epochs - 5000 and b = 50
Precision - 0.95
Accuracy - 0.98
Recall - 1.0

Epochs - 10000 and b = 5
Precision - 0.95
Accuracy - 0.98
Recall - 1.0

Epochs - 10000 and b = 50
Precision - 0.95
Accuracy - 0.98
Recall - 1.0

Epochs - 100 and b = 50
Precision - 0.926829268293
Accuracy - 0.97
Recall - 1.0

Question - 3

Gave :

- **accuracy** = 0.812 on a 7000:4000 of the dataset
- **accuracy** = 0.827 on a 8000:3000 of the dataset

Made a very general version of the decision tree such that there is a binary split at every stage. The attributes having Discrete values were mapped to step integers
{Eg: high - 0, medium - 1, low - 2 } and then the dataset is split based on the value haveing least entropy. Also, each branch should have a minimum of **50** for further division/split of the dataset.
Entropy = -P/T*log(P/T) - N/T*log(N/T)

Question - 4

- K = 5
  Confusion Matrix
  [[1, 0, 0, 0, 0, 0, 0, 0, 0, 0],
  [0, 1, 0, 0, 0, 0, 0, 0, 0, 0],
  [0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
  [0, 0, 1, 0, 0, 0, 0, 0, 0, 1],
  [0, 0, 0, 1, 1, 0, 0, 0, 0, 0],
  [0, 0, 0, 0, 0, 1, 0, 0, 0, 0],
  [0, 0, 0, 0, 0, 0, 1, 0, 0, 0],
  [0, 0, 0, 0, 0, 0, 0, 1, 0, 0],
  [0, 0, 0, 0, 0, 0, 0, 0, 1, 0],
  [0, 0, 0, 0, 0, 0, 0, 0, 0, 0]]

  Accuracy : 0.7
  F_score : 0.6667

- K = 7
  Confusion Matrix
  [[1, 0, 0, 0, 0, 0, 0, 0, 0, 0],
  [0, 1, 0, 0, 0, 0, 0, 0, 0, 0],

[0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
[0, 0, 1, 1, 0, 0, 0, 0, 0, 0],
[0, 0, 0, 0, 1, 0, 0, 0, 0, 0],
[0, 0, 0, 0, 0, 1, 0, 0, 0, 0],
[0, 0, 0, 0, 0, 0, 1, 0, 0, 0],
[0, 0, 0, 0, 0, 0, 0, 1, 0, 0],
[0, 0, 0, 0, 0, 0, 0, 0, 1, 1],
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0]]

Accuracy : 0.8
F_score : 0.7333333333333333

- K = 10

[[1, 0, 0, 0, 0, 0, 0, 0, 0, 0],
[0, 1, 0, 0, 0, 0, 0, 0, 0, 0],
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
[0, 0, 1, 1, 0, 0, 0, 0, 0, 0],
[0, 0, 0, 0, 1, 1, 0, 0, 0, 0],
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0],
[0, 0, 0, 0, 0, 0, 1, 0, 0, 1],
[0, 0, 0, 0, 0, 0, 0, 1, 0, 0],
[0, 0, 0, 0, 0, 0, 0, 0, 1, 0],
[0, 0, 0, 0, 0, 0, 0, 0, 0, 0]]

Accuracy : 0.7
F_score : 0.6