PRML LAB-2

Name: Shalin Jain

Roll No.: - B21CS070

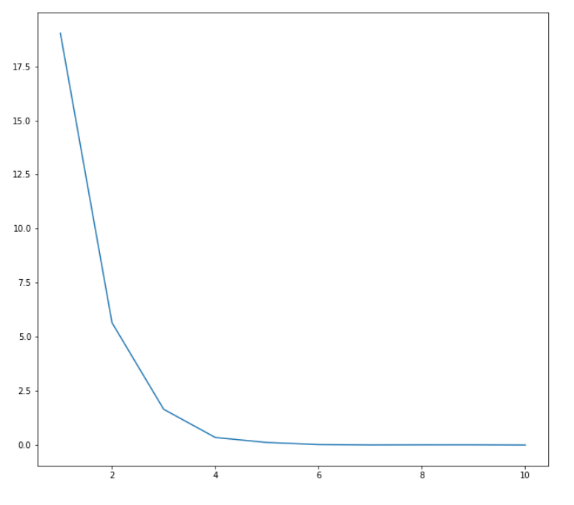
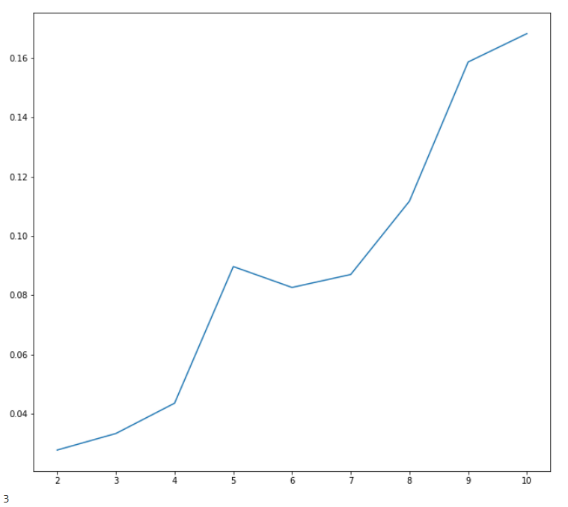
# Problem 1

## Part 1 Pre-processing the data

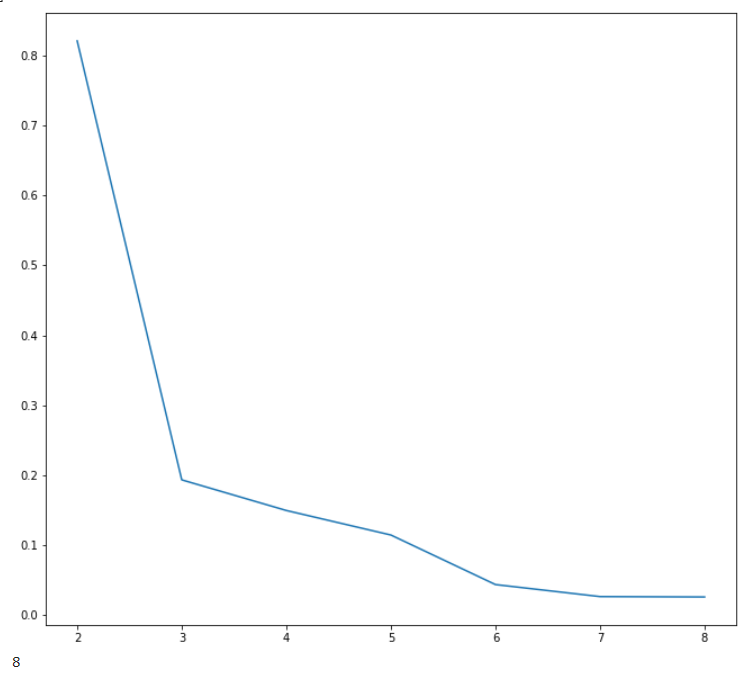
The data was pre-processed by dropping the null values and normalizing the dataset

## Part 2 Varying Hyper-Parameters

We varied three hyper-parameters that are max\_depth, min\_sample\_leaf and max\_features. The obtained graph for mse are as shown

1. Max\_depth ii) Min\_Samples\_Leaf

iii) Max\_Features

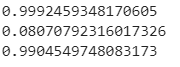


Because max\_depth and min\_sample\_leaf are very useful while defining the height and complexity of the tree we used these hyper-parameters. We took the value of hyper-parameters for which the mse is the lowest.

## Part 3 Performing cross validations

1. Hold out cross validation

Accuracy



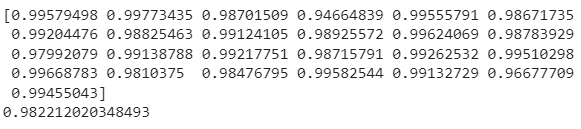
Cross Validation Score

MSE

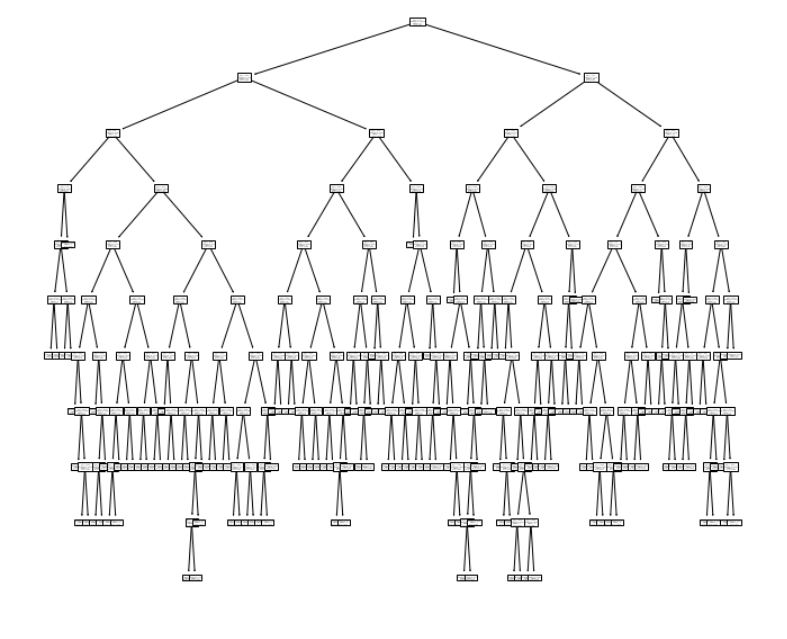
1. 5-Fold Cross Validation



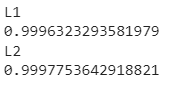
1. Repeated 5-Fold Cross Validation



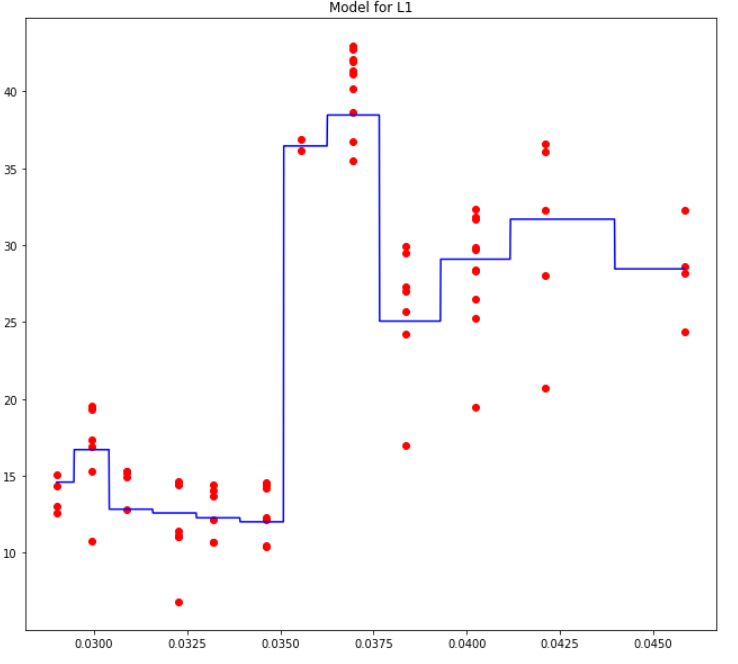
Tree Plot



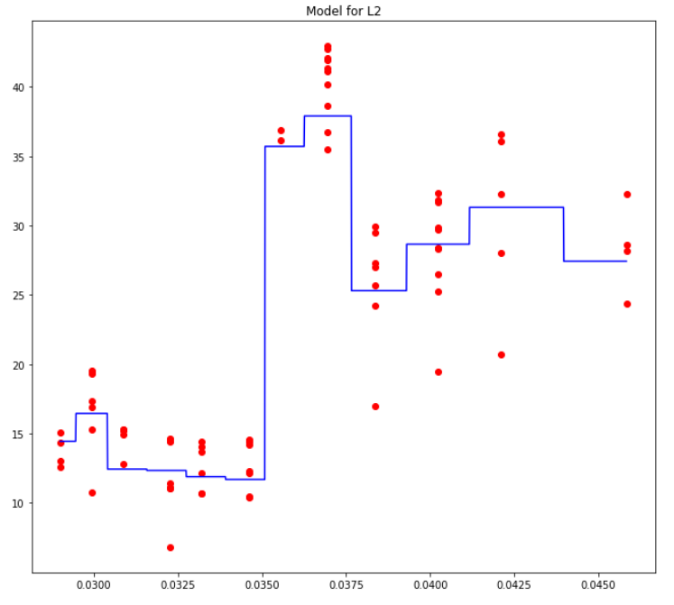
## Part 4 L1 and L2 Loss

The loss function L2 that is mean squared error performs better in case of regression.

1. L1 Loss

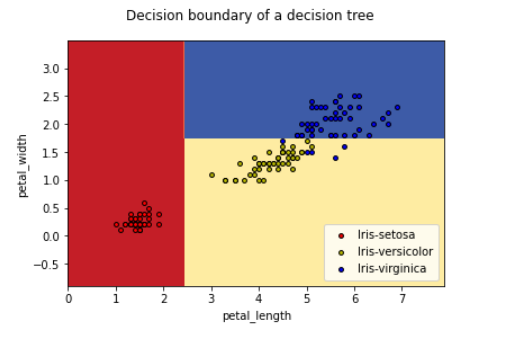


1. L2 Loss

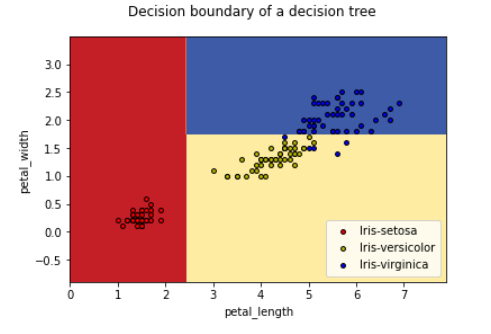


# Problem 2

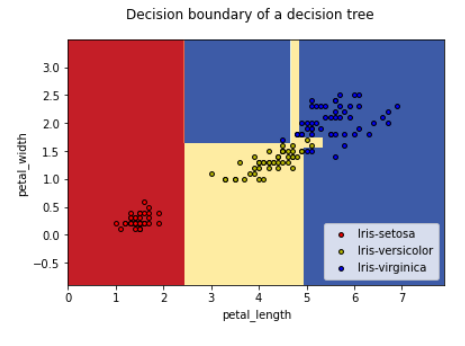
## Part 1 Training a Decision Tree Classifier and Plotting Decision Tree Boundary



## Part 2 Dropping one value and plotting the decision boundary

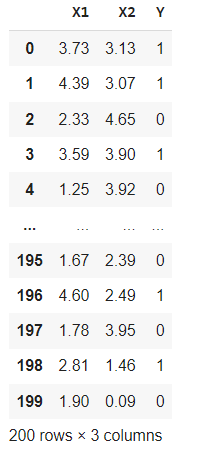


## Part 3 Changing max\_depth to None and plotting decision boundary

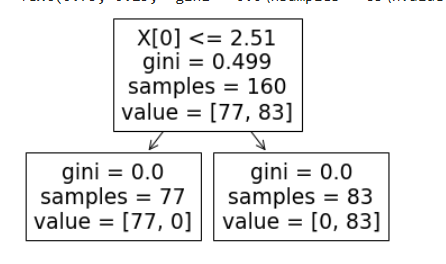


The difference between first and the third plot is that the decision boundary is more precise with increase in max\_depth.

## Part 4 Creating a random dataset



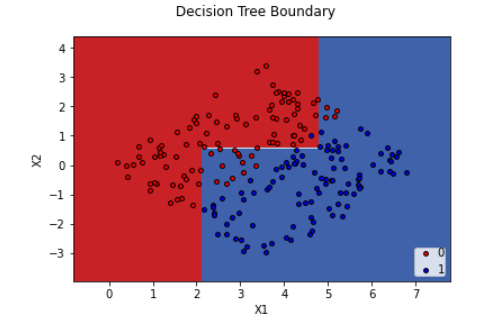
Decision Tree (Max\_depth = 2)



Decision Boundary

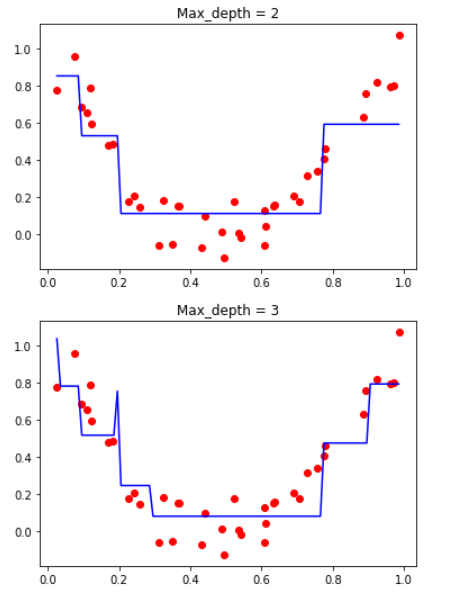


Decision boundary after rotating

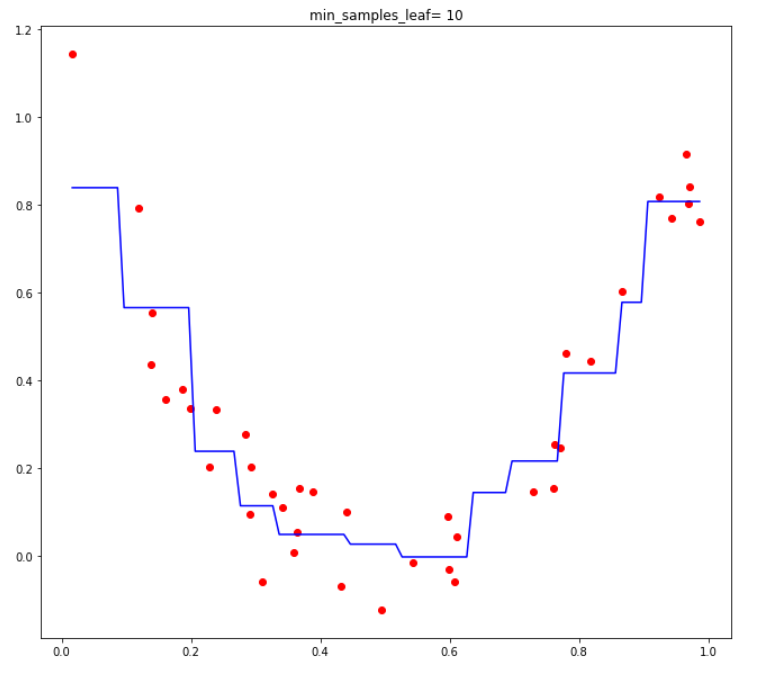


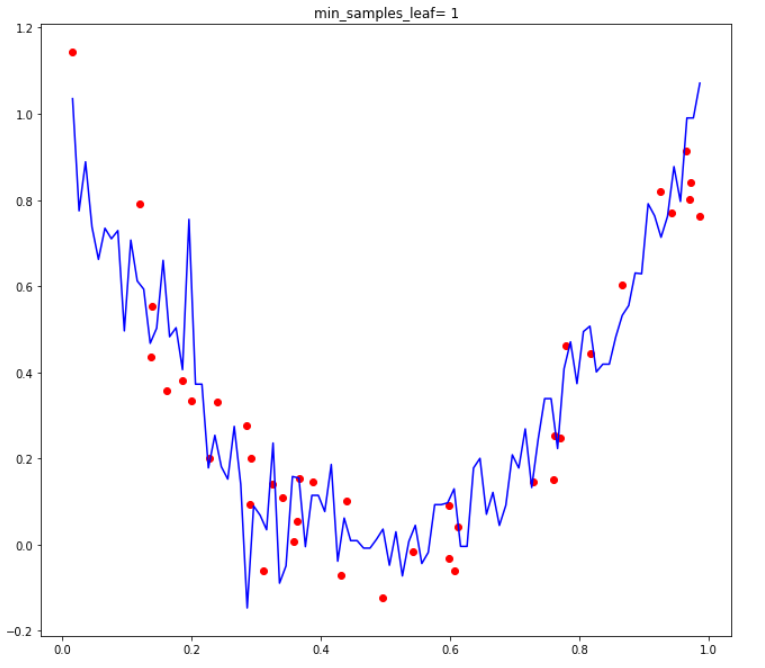
## Regression

### Part 1



### Part 2



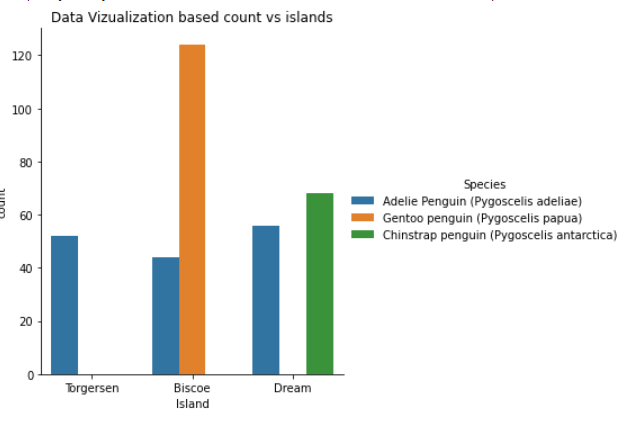


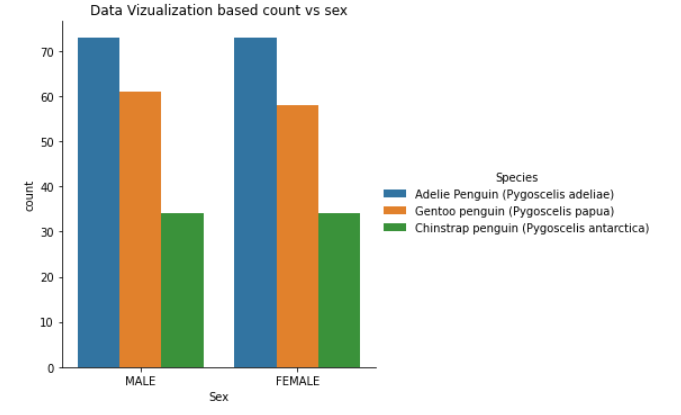
We can clearly see the difference that if minimum samples in a leaf is less then more accurate boundary is present.

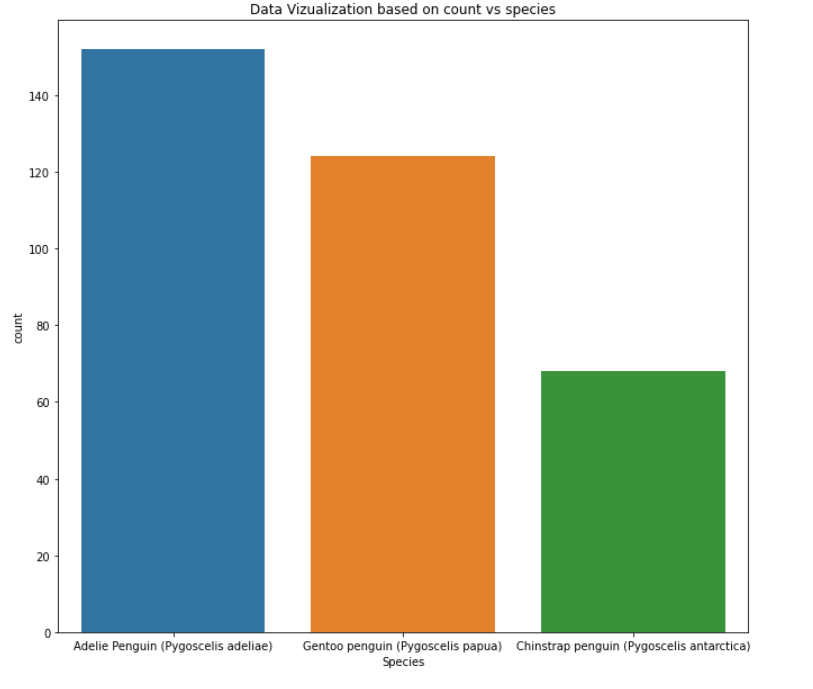
In the part 1 we can decision boundary’s precision and accuracy increases with increase in max\_depth. The increase in max\_depth increase more accurate splits as well as predictions.

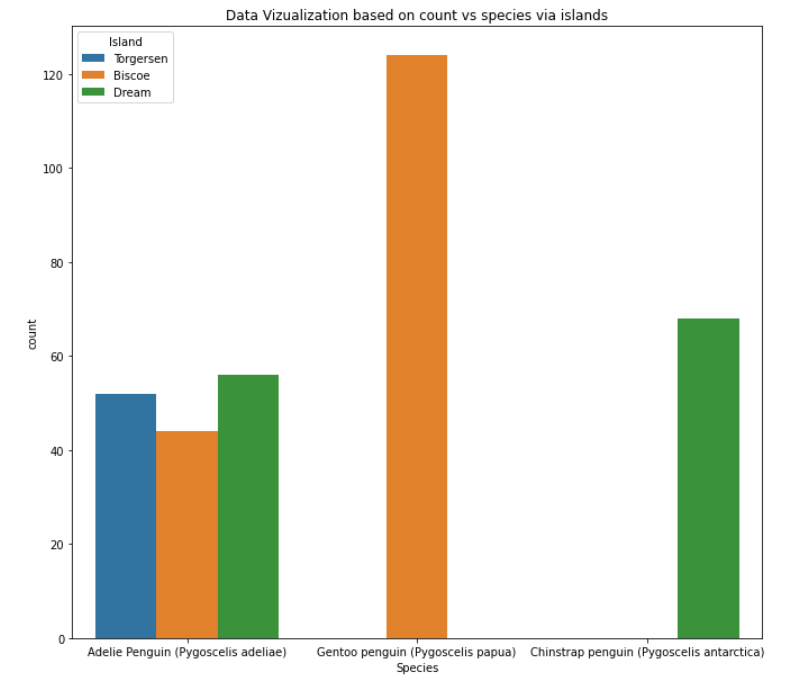
# Problem 3

## Data visualization





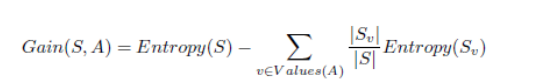




## Decision Tree Class

For the third problem I have used entropy for the loss function and used information gain to get gain for each parent, child attribute and the best gain is used find the best split. The split is calculated with the help of information gain as well as the threshold.

We used the formula given below to find the entropy gain: -



The overall as well as class wise accuracy for my model is as follows

 Where 0,1,2 are the respective classes of species.