Question 3 : Decision Trees

We have performed a grid search on the ***max_depth*** parameter of the decision trees in the range **[1,100)**.

Accuracy Scores for different parameter settings:

a. **Abalone dataset:**
Raw dataset: The best accuracy of 26.26% is obtained at a maximum depth of 4
PCA dataset: The best accuracy of 26.33% is obtained at a maximum depth of 4
LDA dataset: The best accuracy of 25.8% is obtained at a maximum depth of 3

b. **Wine dataset:**
Raw dataset: The best accuracy of 49.97% is obtained at a maximum depth of 4
PCA dataset: The best accuracy of 45.57% is obtained at a maximum depth of 6
LDA dataset: The best accuracy of 53.19% is obtained at a maximum depth of 1

Observations :

Fig 3.1 to 3.6 demonstrate the **Mean Test Accuracy score** vs **Maximum Depth** plots for the 6 datasets. It can be observed from the Accuracy vs Depth plot for all the 6 datasets that the performance of the model first improves with the increase in the maximum depth but degrades rapidly after achieving the peak accuracy score. Thereafter, the accuracy seems to fluctuate in the lower range. This indicates that overfitting may be occurring with the increase in depth. Very low values of depth (in the range of 1-10) seems to be the best choice for all the datasets.

Fig 3.1: Wine - PCA

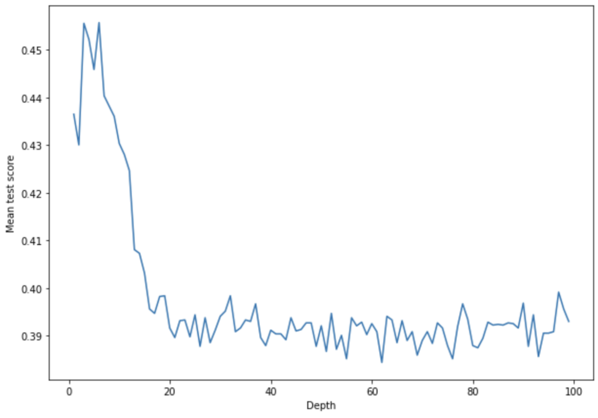


Fig 3.2: Wine-LDA

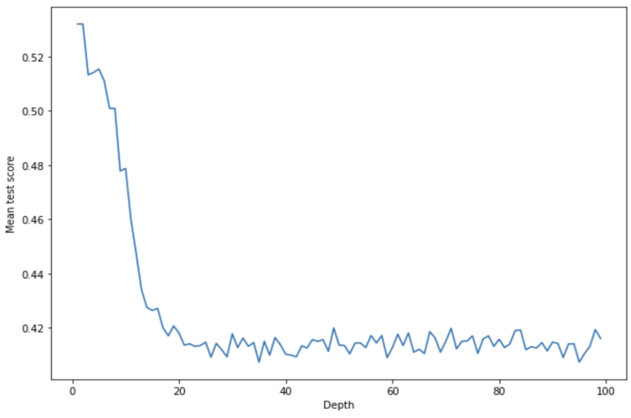


Fig 3.3: Wine-raw

Fig 3.4: Abalone-PCA

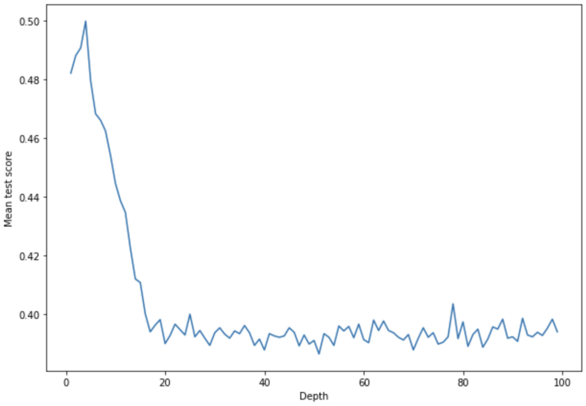


Fig 3.5: Abalone-LDA

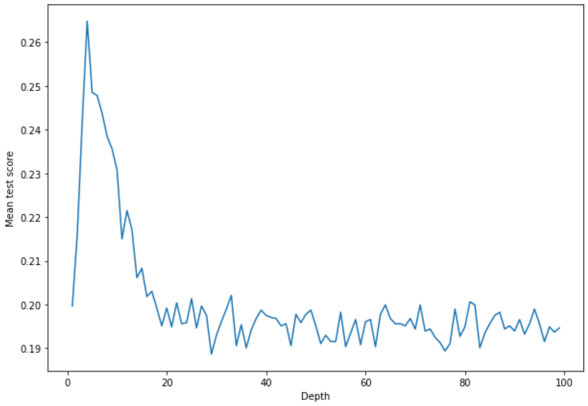
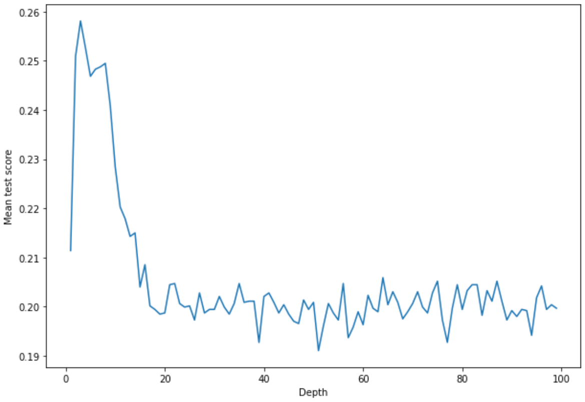
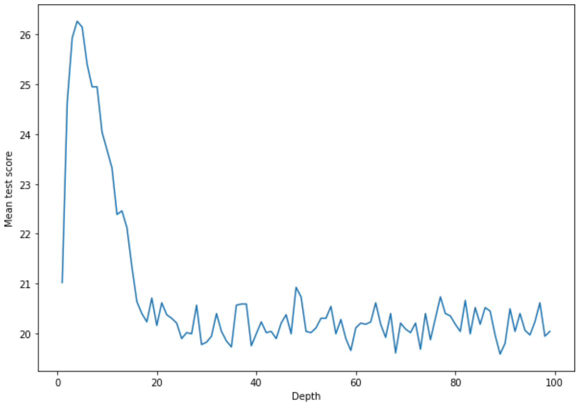


Fig 3.6:Abalone-raw

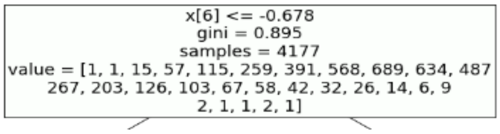




Interpretability:

a. **Abalone dataset:**

Fig 3.5(a) and (b) below show the best decision tree for abalone and its root node respectively. It shows the first partitioning feature used is x[6] i.e ***'Shell Weight'*** which means this feature provided the most information gain and was the most discriminating feature. It should be noted that ***'Shell Weight'*** was also observed to have a high correlation with the target variable i.e ***'Rings'*** as per the analysis performed in Assignment1 and hence corroborates the finding that it contains high information gain. Also, some features like x[2] i.e ***'Height'*** were not included in making any split decision.



x[6] <= -0.678
gini = 0.895
samples = 4177
value = [1, 1, 15, 57, 115, 259, 391, 568, 689, 634, 487
267, 203, 126, 103, 67, 58, 42, 32, 26, 14, 6, 9
2, 1, 1, 2, 1]

Fig 3.7 (b): Root Node(abalone-raw)

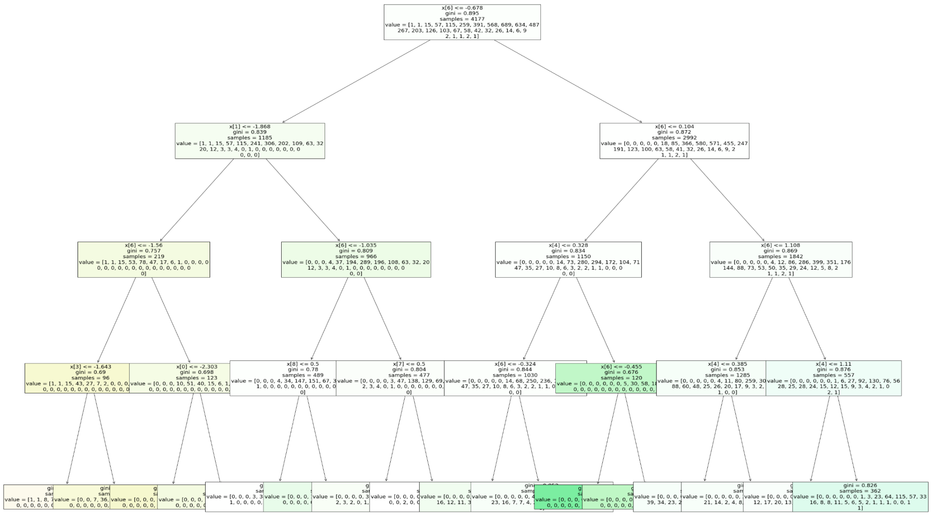

x[6] <= -0.678
gini = 0.895
samples = 4177
value = [1, 1, 15, 57, 115, 259, 391, 568, 689, 634, 487 267, 203, 126, 103, 67, 58, 42, 32, 26, 14, 6, 9 2, 1, 1, 2, 1]
x[1] <= -1.868
gini = 0.839
samples = 1185
value = [1, 1, 15, 57, 115, 241, 306, 202, 109, 63, 32 20, 12, 3, 3, 4, 0, 1, 0, 0, 0, 0, 0, 0, 0 0, 0, 0]
x[6] <= 0.104
gini = 0.872
samples = 2992
value = [0, 0, 0, 0, 0, 18, 85, 366, 580, 571, 455, 247 191, 123, 100, 63, 58, 41, 32, 26, 14, 6, 9, 2 1, 1, 2, 1]
x[6] <= -1.56
gini = 0.757
samples = 219
value = [1, 1, 15, 53, 78, 47, 17, 6, 1, 0, 0, 0, 0 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0 0]
x[6] <= -1.035
gini = 0.809
samples = 966
value = [0, 0, 0, 4, 37, 194, 289, 196, 108, 63, 32, 20 12, 3, 3, 4, 0, 1, 0, 0, 0, 0, 0, 0, 0 0, 0]
x[4] <= 0.328
gini = 0.834
samples = 1150
value = [0, 0, 0, 0, 0, 14, 73, 280, 294, 172, 104, 71 47, 35, 27, 10, 8, 6, 3, 2, 2, 1, 1, 0, 0, 0 0, 0]
x[6] <= 1.108
gini = 0.869
samples = 1842
value = [0, 0, 0, 0, 0, 4, 12, 86, 286, 399, 351, 176 144, 88, 73, 53, 50, 35, 29, 24, 12, 5, 8, 2 1, 1, 2, 1]
x[3] <= -1.643
gini = 0.69
samples = 96
x[0] <= -2.303
gini = 0.698
samples = 123
x[8] <= 0.5
gini = 0.78
samples = 489
x[7] <= 0.5
gini = 0.804
samples = 477
x[6] <= -0.324
gini = 0.844
samples = 1030
x[6] <= -0.455
gini = 0.676
samples = 120
x[4] <= 0.385
gini = 0.853
samples = 1285
x[4] <= 1.11
gini = 0.876
samples = 557
value = [0, 0, 0, 0, 0, 0, 1, 6, 27, 92, 130, 76, 56 28, 25, 28, 24, 15, 12, 15, 9, 3, 4, 2, 1, 0 2, 1]
gini = 0.826
samples = 362
value = [0, 0, 0, 0, 0, 0, 1, 3, 23, 64, 115, 57, 33 16, 8, 8, 11, 5, 6, 5, 2, 1, 1, 1, 0, 0, 1 1]

Fig 3.7(a) : Best performing Decision Tree on Abalone Dataset

b. <u>Wine dataset:</u>

Fig 3.8 (a) and (b) below show the best decision tree for wine and its root node respectively. It shows the partitioning criterion and the feature used to partition it. As we can see, the topmost partition feature is x[10] i.e. ***"Alcohol"*** column in the dataset, which means this feature provided the most information gain and was the most discriminating feature. Another interesting observation is that some features, namely - ***"Citric Acid"***, ***"Total Sulfur Dioxide Density", "pH", "Type"***, weren't even used once in the whole decision tree for splitting, which means those features contributed the least in the decision making process and might be the most irrelevant features. Removal of such features could result in better performance.

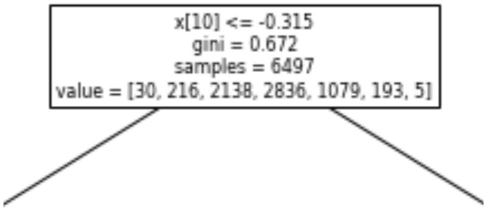


Fig 3.8 (b): Root Node(wine-raw)

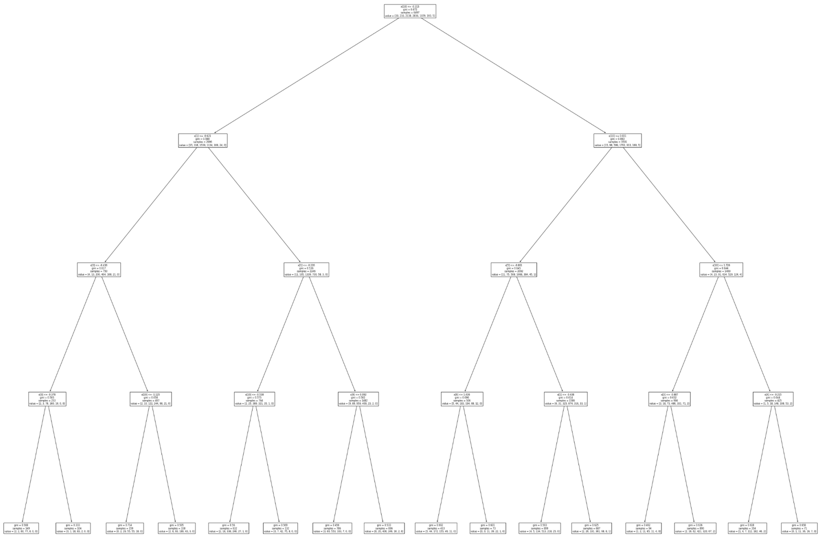

x[10] <= -0.315
gini = 0.672
samples = 6497
value = [30, 216, 2138, 2836, 1079, 193, 5]
x[1] <= -0.621
gini = 0.588
samples = 2996
value = [15, 118, 1539, 1134, 166, 24, 0]
x[10] <= 0.831
gini = 0.663
samples = 3501
value = [15, 98, 599, 1702, 913, 169, 5]
x[0] <= -0.436
gini = 0.617
samples = 750
value = [4, 13, 200, 404, 108, 21, 0]
x[1] <= -0.226
gini = 0.536
samples = 2246
value = [11, 105, 1339, 730, 58, 3, 0]
x[5] <= -0.861
gini = 0.645
samples = 2092
value = [11, 75, 508, 1068, 384, 45, 1]
x[10] <= 1.726
gini = 0.644
samples = 1409
value = [4, 23, 91, 634, 529, 124, 4]
x[9] <= -0.378
gini = 0.503
samples = 253
value = [2, 3, 78, 160, 10, 0, 0]
x[10] <= -1.125
gini = 0.658
samples = 497
value = [2, 10, 122, 244, 98, 21, 0]
x[10] <= -0.538
gini = 0.573
samples = 764
value = [2, 25, 380, 321, 35, 1, 0]
x[9] <= 0.092
gini = 0.582
samples = 1482
value = [9, 89, 959, 400, 23, 2, 0]
x[9] <= 1.426
gini = 0.696
samples = 506
value = [5, 44, 183, 194, 68, 12, 0]
x[1] <= -0.438
gini = 0.614
samples = 1586
value = [6, 31, 325, 874, 316, 33, 1]
x[3] <= -0.887
gini = 0.632
samples = 984
value = [3, 18, 73, 486, 331, 71, 2]
x[4] <= -0.215
gini = 0.648
samples = 425
value = [1, 5, 18, 148, 198, 53, 2]
gini = 0.568
samples = 149
value = [2, 2, 60, 77, 8, 0, 0]
gini = 0.333
samples = 104
value = [0, 1, 18, 83, 2, 0, 0]
gini = 0.714
samples = 159
value = [0, 2, 29, 55, 55, 18, 0]
gini = 0.595
samples = 338
value = [2, 8, 93, 189, 43, 3, 0]
gini = 0.56
samples = 632
value = [2, 18, 338, 246, 27, 1, 0]
gini = 0.569
samples = 132
value = [0, 7, 42, 75, 8, 0, 0]
gini = 0.456
samples = 786
value = [3, 60, 553, 163, 7, 0, 0]
gini = 0.533
samples = 696
value = [6, 20, 406, 246, 16, 2, 0]
gini = 0.692
samples = 433
value = [5, 44, 172, 155, 46, 11, 0]
gini = 0.601
samples = 73
value = [0, 0, 11, 39, 22, 1, 0]
gini = 0.593
samples = 899
value = [4, 5, 134, 513, 218, 25, 0]
gini = 0.625
samples = 687
value = [2, 26, 191, 361, 98, 8, 1]
gini = 0.492
samples = 94
value = [1, 2, 11, 65, 11, 4, 0]
gini = 0.636
samples = 890
value = [2, 16, 62, 421, 320, 67, 2]
gini = 0.618
samples = 354
value = [1, 4, 7, 112, 182, 46, 2]
gini = 0.658
samples = 71
value = [0, 1, 11, 36, 16, 7, 0]

Fig 3.7(a) : Best performing Decision Tree on Wine Dataset