

# Machine Intelligence and Expert Systems

# Computer Assignment-1

## Decision Tree

**Name:** Anand Jhunjhunwala

**Roll Number: 17EC35032**

**Aim:** Train and test a decision tree classifier on the Fisher Iris dataset.

## Results:

### Criterion: Entropy

With Default parameters:

Training Accuracy: 100%

Test Accuracy: 96.67%

### Results after variation in parameters Max\_depth and min\_samples\_leaf

- In the table, for each max\_depth value the first row is training accuracy and the next row is test accuracy in %.

[illegible]

## Criterion: Gini

With Default parameters:

Training Accuracy: 100%

Test Accuracy: 96.67%

Results after variation in parameters Max\_depth and min\_samples\_leaf

- In the table, for each max\_depth value the first row is training accuracy and the next row is test accuracy in %.

Max_depth	Min_samples_leaf = 1	Min_samples_leaf = 5	Min_samples_leaf = 8	Min_samples_leaf = 11	Min_samples_leaf = 15	Min_samples_leaf = 20	Min_samples_leaf = 30	Min_samples_leaf = 40	Min_samples_leaf = 50
None	100	96.67	96.67	96.67	96.67	96.67	96.67	66.67	66.67
	96.67	93.33	93.33	93.33	93.33	93.33	93.33	66.67	66.67
4	99.17	96.67	96.67	96.67	96.67	96.67	96.67	66.67	66.67
	93.33	93.33	93.33	93.33	93.33	93.33	93.33	66.67	66.67
3	97.50	96.67	96.67	96.67	96.67	96.67	96.67	66.67	66.67
	90.00	93.33	93.33	93.33	93.33	93.33	93.33	66.67	66.67
2	96.67	96.67	96.67	96.67	96.67	96.67	96.67	66.67	66.67
	93.33	93.33	93.33	93.33	93.33	93.33	93.33	66.67	66.67
1	66.67	66.67	66.67	66.67	66.67	66.67	66.67	66.67	66.67
	66.67	66.67	66.67	66.67	66.67	66.67	66.67	66.67	66.67

---

## Conclusion:

- With default parameters for both entropy and gini, I got 100% train accuracy indicating overfitting.
- To avoid overfitting, I varied the max\_depth and min\_samples\_leaf parameters for both the cases and In both cases the best accuracy without overfitting is:
  - Training Accuracy: 96.67%
  - Test Accuracy: 93.33%
- Since the data contain 4 different parameters which can act as splitting nodes so max\_depth is varied from 4 to 1 and no other value is checked.
- Since Training examples contain 120 examples so upto approx 30 min\_samples\_leaf I don't observe any difference in accuracy and after that accuracy starts decreasing.