Department of Mathematical and Computational Sciences
National Institute of Technology Karnataka, Surathkal

https://sam.nitk.ac.in/ sam@nitk.edu.in

MA722 - Foundations of Machine Learning Algorithms

Lab problem sheet 6

---

1. For the "Play Tennis" dataset shown below, compute the entropy of the target variable (*PlayTennis*). Then, for each feature (*Outlook*, *Temperature*, *Humidity*, *Wind*), calculate the Information Gain and determine which feature provides the best split for the first decision.

Outlook	Temperature	Humidity	Wind	PlayTennis
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Strong	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Weak	Yes

2. Using the same dataset in Problem 1, construct a simple Decision Tree (without using scikit-learn) using the computed Information Gain values. Display the structure of the Decision Tree in a text-based or flowchart format, showing how decisions are made at each split.

3. Using `scikit-learn`, build a Decision Tree Classifier for the "Weather and Picnic" dataset to predict whether people will go on a picnic (*Picnic*) based on weather conditions (*Outlook*, *Temperature*, *Humidity*, *Wind*). Visualize the resulting Decision Tree as a flowchart and identify the root node chosen by the model.

Outlook	Temperature	Humidity	Wind	Picnic
Sunny	Hot	High	Weak	No
Sunny	Mild	High	Strong	No
Overcast	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Mild	High	Strong	Yes
Rain	Mild	High	Weak	Yes
Sunny	Cool	Normal	Weak	Yes
Overcast	Hot	Normal	Strong	Yes
Sunny	Mild	Normal	Weak	Yes

4. Given the dataset below containing house attributes (*Area*, *Rooms*, and *Price*), calculate the Mean Squared Error (MSE) for different possible splits manually. Identify which feature and threshold value yield the minimum MSE, and thus form the best split for a regression decision tree.

Area (sqft)	**Rooms**	**Price (in $1000)**
850	2	100
900	3	120
1000	3	150
1200	4	200
1500	4	250
1700	5	275
2000	5	310

5. Using `scikit-learn`, build a Decision Tree Regressor to predict the price of cars based on *Engine Size* and *Horsepower*. Train the model, visualize the regression tree, and use it to predict the price for a car with *Engine Size* = 2.5 and *Horsepower* = 180.

Engine Size (L)	**Horsepower**	**Price (in $1000)**
1.2	70	12
1.5	90	15
1.8	110	20
2.0	130	25
2.2	150	28
2.5	170	35
3.0	200	45
3.5	250	60
4.0	300	70

*****