
Tutorial Sheet 8: Advanced Techniques

Course: CSEG 2036P | *School of Computing Sciences, UPES*

Faculty: Dr. Mrityunjay Guha Majumdar

1. Let us say we have to predict whether to play tennis based on weather conditions. We are given the following data

Outlook	Temperature	Humidity	Wind	Play Tennis
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rainy	Mild	High	Weak	Yes
Rainy	Cool	Normal	Weak	Yes
Rainy	Cool	Normal	Strong	No
Overcast	Cool	Normal	Strong	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rainy	Mild	Normal	Weak	Yes

Table 1: Data for Suitability of Weather for Tennis

Use the Information Gain and Entropy approach for Decision Trees for node allocation and devising a flow-chart to decide on weather suitability of tennis.

2. Let us say we are interested in predicting whether a person has COVID based on symptoms and travel history, with the following data given to us:

Fever	Cough	Shortness of Breath	Travel History	COVID
Yes	Yes	No	Yes	Yes
No	Yes	Yes	No	No
Yes	No	No	Yes	Yes
No	Yes	Yes	Yes	Yes
Yes	No	Yes	No	No
No	No	No	No	No
Yes	Yes	Yes	Yes	Yes
No	No	No	No	No
Yes	Yes	No	No	No
No	No	Yes	Yes	Yes

Table 2: Data for Predicting about COVID

Use the Gini Index approach for Decision Trees for node allocation and devising a flow-chart to decide on predictability of COVID using symptoms and travel history.

3. Suppose you have a neural network with one input layer consisting of three features (X_1, X_2, X_3) , one hidden layer with two neurons, and an output layer. The activation function used in the hidden layer is the sigmoid function $\sigma(x) = \frac{1}{1+e^{-x}}$. If you provide an input vector $(1, 0, 1)$, what are the steps involved in calculating the final output of the neural network?

Assume $w_{11} = 0.5, w_{21} = -0.3, w_{31} = 0.8, b_1 = 0.2, w_{12} = -0.4, w_{22} = 0.7, w_{32} = 0.1, b_2 = -0.5, w_1 = 0.6, w_2 = -0.2, b_{\text{output}} = 0.1$, where w_{ij} is the weight connecting the i^{th} neuron in the input layer to the j^{th} neuron in the hidden layer, b_j is the bias for the j^{th} neuron in the hidden layer, a_j is the output (activation) of the j^{th} neuron in the hidden layer, w_k is the weight connecting the k^{th} neuron in the hidden layer to the output neuron, b_{output} is the bias for the output neuron.

4. *You are given a dataset containing information about smartphone users, including features such as screen size (in inches), battery life (in hours), and average daily app usage time (in minutes). The dataset includes the following observations:

User ID	Screen Size (in inches)	Battery Life (in hours)	App Usage (in minutes)
1	5.0	12	120
2	6.2	18	180
3	5.5	15	150
4	4.7	10	90
5	6.5	20	200
6	5.2	14	160
7	4.0	8	60
8	6.0	16	170

Table 3: Data for Smartphone Market Segmentation

Perform a k-means clustering analysis on this dataset to group users based on their smartphone characteristics. Choose an appropriate number of clusters for segmentation. Discuss the results and potential insights gained from the clustering. Additionally, propose one business strategy that could be implemented based on the identified clusters.