Quiz 1

Total time: 7 mins (10:05 AM to 10:12 AM)

1. Let us work with an image dataset. Let us call is 1024-ImageNet. It has images of 1024 categories of objects -- like: child, fish, stone, rain, etc. Given any dataset of 1024-ImageNet, what is the maximum entropy? Explain. [1]

2. Should we stop growing a decision tree if there is no reduction in entropy?. Explain with example or counter-example. [1]

3. Given the following dataset and attributes: Day, Outlook, Temperature, Humidity, Windy -- what attribute will be chosen as the root node of the decision tree? Why? [1]

Training Data

Day	Outlook	Temp	Humidity	Windy	Play
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

4. Entropy is one statistical measure to encode the degree of randomness or impurity of a dataset. Can you come up with some other metric that can encode degree of randomness or impurity. Explain with an example. [2]