Tutorial 10:
ID3 Decision Trees and Naïve Bayes Classification

Name	Student no.

For this tutorial, you can work in groups of 1 or 2. Submit the answers to each of the 2 Questions.

Question 1

Consider the following data set (from Mitchell, *Machine Learning*, 1997).

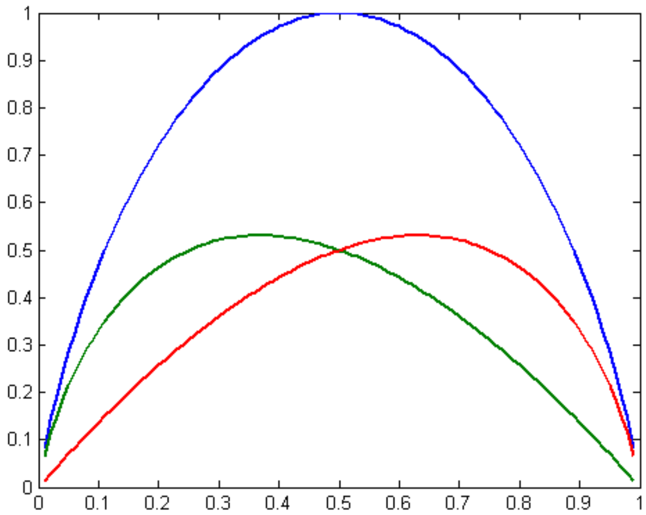
Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes

$$Gain(A) = I\left(\frac{p}{p+n}, \frac{n}{p+n}\right) - Remainder(A)$$

$$Remainder(A) = \sum_{i=1}^{v} \frac{p_i + n_i}{p+n} I\left(\frac{p_i}{p_i+n_i}, \frac{n_i}{p_i+n_i}\right)$$

$$I(P(v_1),\dots,P(v_n)) = \sum_{i=1}^{n} -P(v_i)\log_2 P(v_i)$$

$$I\left(\frac{p}{p+n}, \frac{n}{p+n}\right) = -\frac{p}{p+n}\log_2\frac{p}{p+n} - \frac{n}{p+n}\log_2\frac{n}{p+n}$$



a) Using ID3 (maximum entropy to choose top attribute), determine the decision tree from D1-D10 and draw the tree.

b) Using the tree you just created, count the number of Yes and No under each leaf node and label it with the majority outcome.

c) Using your labels check the outcome of the following test cases (D11-D14).

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D11	Sunny	Mild	Normal	Strong	(Yes)
D12	Overcast	Mild	High	Strong	(Yes)
D13	Overcast	Hot	Normal	Weak	(Yes)
D14	Rain	Mild	High	Strong	(No)

d) Which attribute seems to be critical for handling the failing test case(s)?

Question 2

Construct a Naive Bayes model for the same data as Question 2 (D1-D10) and apply it to the same test data (D11-D14), detailing the model and the test results.

What you need to know: (reviewing theory)

The Naive Bayes assumption is that the joint probability of the attributes (e.g. P(*x*new|*c*new,d) can be decomposed into the product of the individual attribute probabilities. The resulting classification rule is as follows:

$$P(c_{new} \mid x_{new}, d) \propto P(x_{new} \mid c_{new}, d) P(c_{new} \mid d) \propto P(c_{new} \mid d) \prod_{i=1}^{m} P(x_{i,new} \mid c_{new}, d)$$

*x*new is a new input vector, *c*new is a possible class (output) for this input, d is the dataset (D1-D10). The P(*c*new|*x*new,d) values must sum to 1 over the possible classes *c*new. You can do this renormalisation once you have calculated the un-normalised values for each of the (2) output classes.