2. This question is concerned with **decision trees**. Consider the following dataset which relates to whether a person plays golf based on the weather conditions.

Temperature	**Humidity**	**Windy**	**Play**
hot	high	false	no
hot	high	true	no
hot	high	false	yes
mild	high	false	yes
cool	normal	false	yes
cool	normal	true	no
cool	normal	true	yes
mild	high	false	no
cool	normal	false	yes
mild	normal	false	yes
mild	normal	true	yes
mild	high	true	yes
hot	normal	false	yes
mild	high	true	no

a) Write down the equation for the **entropy** $E(p_1, \ldots, p_k)$ of a probability distribution $(p_1, \ldots, p_k)$.

(2 marks)

b) Compute the **information gain** of splitting the data objects using the '*Temperature*' and '*Humidity*' attributes respectively for the weather dataset assuming that '*Play*' represents the class. Show the details of your working. Which attribute of the two is the better one to select for a decision tree?

(14 marks)

c) What is the purpose of pruning a decision tree? Explain what is meant by each of **pre-pruning** and **post-pruning** of a decision tree. What is the main advantage of post-pruning over pre-pruning?

(9 marks)