1. We run the ID3 algorithm for learning decision trees over a set of attributes, where each attribute can take two values. Assume that we have to choose one attribute out of a possible four attributes, which split the data into two groups of 600 data points each, and the positive examples are also 600 in total. We show the distribution of one group for four attributes as following. Which of the attributes will ID3 choose?

Options:

- 100 positive examples, 500 negative examples
- 200 positive examples, 400 negative examples
- 300 positive examples, 300 negative examples
- 400 positive examples, 200 negative examples

**Ans:** 100 positive examples, 500 negative examples

2. What is the correct computation for the entropy of a dataset with 10 positive examples and 40 negative examples?

Options:

- $-\frac{10}{50} \log\left(\frac{10}{50}\right) - \frac{40}{50} \log\left(\frac{40}{50}\right)$
- $-\frac{10}{50} \log\left(\frac{40}{50}\right) - \frac{40}{50} \log\left(\frac{10}{50}\right)$
- $-\frac{10}{50} \log\left(\frac{10}{50}\right) - \frac{40}{10} \log\left(\frac{40}{10}\right)$
- $-\frac{10}{50} \log\left(\frac{10}{50}\right) - \frac{10}{50} \log\left(\frac{40}{50}\right)$

**Ans:** $-\frac{10}{50} \log\left(\frac{10}{50}\right) - \frac{40}{50} \log\left(\frac{40}{50}\right)$

3. What kinds of boolean functions can decision trees represent?

Options:

- All Boolean functions
- Conjunctions
- Disjunctions
- Conjunctions and Disjunctions

**Ans:** All Boolean functions

4. Consider two different approaches for the same learning problem. In which of these two scenarios are you more likely to over fit the training data?

Options:
• Learning decision trees where the depth of the learned trees can be at maximum 10.
• Learning decision trees where their depth is not limited in any way.

Ans: Learning decision trees where their depth is not limited in any way.

5. In a machine learning task that takes a pair of character strings composed of letters in a, b, c, z as examples, we define two feature types:

• The first character in the first string
• The first character in the second string

The dimensionally of the feature space will then be:

Options:

• 52
• 2
• 26
• 54

Ans: 52

6. In a machine learning task that takes a pair of character strings composed of letters in a, b, c, z as examples, we define two feature types:

• The first character in the first string
• The first character in the second string

If features in each examples are sorted by alphabetic order, and the features corresponding to type (i) appear before those that correspond to type (ii), which of the following examples is the feature based representation of Bill Clinton?

Options:

• 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0; 0, 0, 1, 0, 0, 0,
  0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0
• 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, 0, 0; 0, 0, 1, 0, 0,
  0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0
• 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0
• 1, 1

Ans: 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0; 0, 0, 1, 0, 0, 0,
  0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0