

Assignment A5

ST228

13 February 2025

1. Information theory

- (a) Which of the following is/are true about entropy $H(P)$ of a distribution P
 - i. is always defined in bits
 - ii. $H(P) > 0$
 - iii. is the expected value of the information content of a distribution
 - iv. $H(P) \leq 1$
 - v. is defined for continuous distributions too
- (b) In the class, we defined what Kullback-Leibler divergence $D(P\|Q)$ is. What is the condition on Q for $D(P\|Q)$ to be well defined?
- (c) Consider two distributions P and Q as shown below.

	$X = x_1$	$X = x_2$
P	0.3	0.7
Q	0.5	0.5

Table 1: Sample Table

Compute $D(P\|Q)$ and verify that cross-entropy is higher than entropy for in either direction.

2. Decision trees

- (a) Which of the following is/are true about decision trees
 - i. There are only binary trees
 - ii. We cannot mix both categorical and numerical data while building trees
 - iii. Trees can be imbalanced
 - iv. Number of leaves in the tree equals the number of rows in the training dataset
- (b) Suppose we are building a decision tree for a dataset consisting of N datapoints with k ($N \gg k$) binary attributes. What is the maximal number of leaf nodes that this decision tree can have?
 - i. N
 - ii. 2^k
 - iii. $2^k - 1$
 - iv. kN
- (c) Consider the following decision tree:

By convention, each non-leaf node corresponds to a question asked on a feature. The number mentioned in the leaf nodes are the number of entries in the training data that corresponds to a particular leaf node. Assuming the real-world data has the same distribution as the training dataset, calculate the average number of questions asked.

