1 Hypothesis Space

Suppose we have an instance space consisting of 4 features X_1, X_2, X_3, X_4 and a label y such that y is determined by a function of x, $y = f(X_1, X_2, X_3, X_4)$. X_1 and X_2 can take 3 different values, while X_3 and X_4 can take 4 different values. The label y can take in $\{0, 1\}$.

- What is the number of possible functions f?
- Suppose we have a training algorithm that generates classifiers that use only one feature. What is the size of the hypothesis space searched?
- Suppose we have a training algorithm that generates classifiers that use only two features. What is the size of the hypothesis space searched?

2 Decision Tree

2.1 Question 1

We will use the dataset below to learn a decision tree which predicts the class label which is binary {T,F}.

ID	A	B	C	Class
1	1	0	0	T
2	1	0	1	F
3	0	1	0	F
4	1	1	1	F
5	1	1	0	T

- What is the entropy of the root node? Show your working.
- Draw the complete decision tree.
- Are there feature(s) that are redundant?

2.2 Question 2

Consider the matrix below.

$$M = \begin{bmatrix} 3 & 5 & 8 & 1 & 6 & 4 \\ 2 & 7 & 6 & 3 & 4 & 5 \\ 5 & 7 & 9 & 2 & 8 & 6 \end{bmatrix} \tag{1}$$

where M^T represents represents 6 sample points, each with three features f_1 f_2, and f_3.
The labels for the data are

$$\begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}^T \tag{2}$$

We would like to build a decision tree of depth 2 by hand to classify the data.

- What is the entropy of the root node? Show your working.
- What is the rule for the first split? (Write it in a rule format for example, $f_1 > 2$.) Show your working.
- What is the rule for the second split? Show your working.
- Should you have a third split, explain your answer?