FMCS1 Lab Session 5

Manuel Marques-Pita and Frank Keller

March 22, 2005

1 Introduction

During this lab session you will be calculating the entropy of a data set and devising codes to make the space it takes (for storage) smaller. You will also help the evolution of a Saturnian frog by calculating Mutual Information measures. (Answer all your questions in a MatLab .m file)

2 Task 1: Basic Information Theory

We have a large collection of of items to be organised in an inventory. Each item is marked with a label and a number the labels are `A,B, C` and `D` and the numbers are just positive integers used to number the items. We have 512 items of type A, 256 of type B, 128 of type C and 128 of type D.

For our inventory, we initially assigned the following codes

```
A --> 00
B --> 01
C --> 10
D --> 11
```

Question 1 How many bits are necessary in order to encode the inventory?

Question 2 If we assigned different codes of different length to the items, can you devise a way of assigning these codes which would make the inventory shorter?

Question 3 Calulate the entropy $H(X)$ for this set. Can you think of a way of using this number in order to determine how small this inventory can be?

3 Task 1: Mutual Information

We are now in *Enceladus*, one of Saturn's moons which, scientists have just found out, has an atmosphere! In that distant frozen world, we are going to imagine the existence of some life form that is evolving through learning how to survive. We asked our friends in the Cassini mission to send some data on this frog-looking creature, specifically about the decisions it makes when food is available (whether to eat it or not). The summary of these data is shown in Figure 1.

Example	Attributes					
Examples	Hungry	Competition	Danger	Food Value	Help Available	Goal
E1	Yes	High	No	High	Yes	Yes
E2	Yes	High	Yes	Medium	No	No
E3	No	High	No	High	No	No
E4	No	Low	No	High	No	Yes
E5	Yes	High	Yes	High	Yes	Yes
E6	Yes	Medium	Yes	Medium	Yes	Yes
E7	Yes	Low	No	Low	Yes	Yes
E8	No	Medium	No	Low	No	No
E9	No	Medium	Yes	Medium	Yes	No
E10	No	None	Yes	Medium	No	Yes

Figure 1: Summary of situations in which eating will or will not be attempted

Question 4 Calculate the Entropy for the set of decisions made by the Enceladonian frog regarding whether or not to eat some available food.

Now we want to understand how the Enceladonian frog could learn what questions to ask first, in order to make a decision more quickly! If, for example, we use the attribute *Hungry* to split the set of data, we will have two branches (for the two values of this attribute, *yes* and *no*) for each of these branches, some examples will be classified as positive (for example branch:YES, E1:Yes) and some as negative (like E2). Figure 2 shows the split for the attribute *Hungry* in this example. Every time we split a data set in this way, we can calculate the mutual information between the set and the split. This could help the

Enceladonian frog to make decisions more quickly in the future!

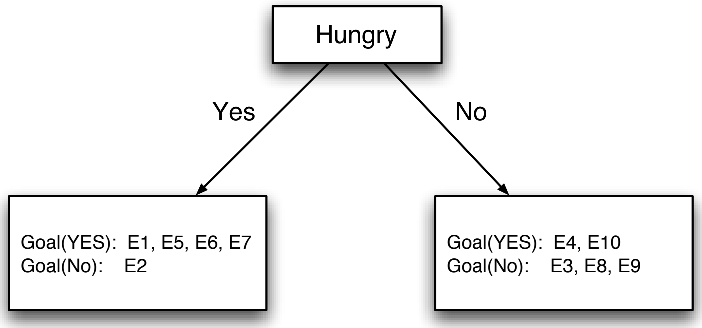

Figure 2: Split for the attribute *Hungry* in this example

The formula to calculate the Mutual Information is the following

$I(X;Y) = H(X) - H(X|Y)$

Where $H(X)$ is the standard Entropy measure for the whole set (calculated previously) and

$H(X|Y) = \sum_x p(x)H(Y|X = x)$

Which is the Entropy of the set given the split on attribute Y

Question 5 Calculate the mutual information between the set and the partitions for the attributes *Hungry, Competition* and *Food Value* and decide which one is the best attribute to ask for when trying to make a decision for the Enceladonian frog in this context.

Important Note. All the answers to these questions have to be submitted by the end of the lab in an .m (MatLab) file. send them to `m.marques-pita@ed.ac.uk`