Homework #3: Machine Learning

Kyler Little

February 26, 2018

Problem #1

Suppose that we have three coloured boxes r (red), b (blue), and g (green). Box r contains 3 apples, 4 oranges, and 3 limes; box b contains 1 apple, 1 orange, and 0 limes; and box g contains 3 apples, 3 oranges, and 4 limes. If a box is chosen at random with probabilities $p(r) = 0.2$, $p(b) = 0.2$, $p(g) = 0.6$, and a piece of fruit is removed from the box (with equal probability of selecting any of the items in the box), then what is the probability of selecting an apple? If we observe that the selected fruit is in fact an orange, what is the probability that it came from the green box?

Problem #2

Given the following data set containing three attributes and one class, use nave Bayes classifier to determine the class (Yes/No) of Stolen for a Red Domestic SUV.

Example No.	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes

Problem #3

This question is about nave Bayes classifier. Please do the following:
(a) State what is the simplifying assumption made by nave Bayes classifier.
(b) Given a binary-class classification problem in which the class labels are binary, the dimension of feature is d, and each attribute can take k different values. Please provide the numbers of parameters to be estimated with AND without the simplifying assumption. Briefly justify why the simplifying assumption is necessary.

Problem #4

Assume we want to classify science texts into three categories physics, biology and chemistry. The following probabilities have been estimated from analyzing a corpus of pre-classified web-pages gathered from Yahoo.
Assuming ...

Problem #5

Consider the following table of observations:
From the classified examples in the above table, construct two decision trees (by hand) for the classification "Play Golf." For the first tree, use Temperature as the root node. (This is a really bad choice.) Continue the construction of tree as discussed in class for the subsequent nodes using information gain. Remember that different attributes can be used in different branches on a given level of the tree. For the second tree, follow the Decision Tree Learning algorithm described in class. At each step, choose the attribute with the highest information gain. Work out the computations of information gain by hand and draw the decision tree.