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?
In regards to the first question, we need to compute the marginal probability $p(\text{apple})$. We can do this using the sum and product rules. Let Y denote coloured box.

$$\begin{aligned}
p(\text{apple}) &= \sum_Y p(\text{apple}, Y) = \sum_Y p(\text{apple}|Y)p(Y) \\
&= p(\text{apple}|r)p(r) + p(\text{apple}|b)p(b) + p(\text{apple}|g)p(g) \\
&= 0.3(0.2) + 0.5(0.2) + 0.3(0.6) \\
&= 0.34
\end{aligned}$$

The second question asks us to compute $p(g|\text{orange})$. We have all of the conditional properties of the form $P(\text{fruit}|\text{box})$, and we want $P(\text{box}|\text{fruit})$. This is exactly the purpose of Bayes' Theorem, and so we apply it.

$$\begin{aligned}
p(g|\text{orange}) &= \frac{p(\text{orange}|g) * p(g)}{p(\text{orange})} \\
&= \frac{p(\text{orange}|g) * p(g)}{p(\text{orange}|r)p(r) + p(\text{orange}|b)p(b) + p(\text{orange}|g)p(g)} \\
&= \frac{0.3(0.6)}{0.4(0.2) + 0.5(0.2) + 0.3(0.6)} \\
&= 0.5
\end{aligned}$$

Problem #2

Given the following data set containing three attributes and one class, use naïve 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

Let $x^* = [\text{Red SUV Domestic}]^T$. To use naïve Bayes' classifier, we shall calculate $P(\text{Yes} \,|\, x^*)$ and $P(\text{No} \,|\, x^*)$. Whichever is larger will be the decision. To start, we will calculate $P(\text{Yes} \,|\, x^*)$. We ignore the denominators because they are the same, and we only care which is larger.

$$\begin{aligned}
P(\text{Yes} \,|\, x^*) &\propto P(x^* \,|\, \text{Yes}) * P(\text{Yes}) \\
&\propto (P(\text{Red} \,|\, \text{Yes}) * P(\text{SUV} \,|\, \text{Yes}) * P(\text{Domestic} \,|\, \text{Yes})) * P(\text{Yes}) \\
&\propto ((3/5) * (1/5) * (2/5)) * 0.5 \\
&\propto 0.024
\end{aligned}$$

Next, we calculate $P(\text{No} \,|\, x^*)$, again ignoring the denominator.

$$\begin{aligned}
P(\text{No} \,|\, x^*) &\propto P(x^* \,|\, \text{No}) * P(\text{No}) \\
&\propto (P(\text{Red} \,|\, \text{No}) * P(\text{SUV} \,|\, \text{No}) * P(\text{Domestic} \,|\, \text{No})) * P(\text{No}) \\
&\propto ((2/5) * (3/5) * (3/5)) * 0.5 \\
&\propto 0.072
\end{aligned}$$

Since $P(\text{No} \,|\, x^*) > P(\text{Yes} \,|\, x^*)$, our decision is "No."

Problem #3

This question is about naïve Bayes classifier. Please do the following:
(a) State what is the simplifying assumption made by naïve 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.

c	Physics	Biology	Chemistry
P(c)	Red	Sports	Domestic
P(atom | c)	Red	Sports	Domestic
P(carbon | c)	Red	Sports	Domestic
P(proton | c)	Yellow	Sports	Domestic
P(life | c)	Yellow	Sports	Imported
P(earth | c)	Yellow	SUV	Imported

Assuming that the probability of each evidence word is independent of other word occurrences given the category of the text, compute the (posterior) probability for each of the possible categories each of the following short texts; and based on that, their most likely classification. Assume that the categories are disjoint and exhaustive (i.e., every text is either physics, or biology or chemistry, and no text can be more than one). Assume that words are first stemmed to reduce them to their base form (atoms → atom) and ignore any words that are not in the table:

A: the carbon atom is the foundation of life on earth
B: the carbon atom contains 12 protons.

Problem #5

Consider the following table of observations:

No.	Outlook	Temperature	Humidity	Windy	Play Golf?
1	sunny	hot	high	false	N
2	sunny	hot	high	true	N
3	overcast	hot	high	false	Y
4	rain	mild	high	false	Y
5	rain	cool	normal	false	Y
6	rain	cool	normal	true	N
7	overcast	cool	normal	true	Y
8	sunny	mild	high	false	N
9	sunny	cool	normal	false	Y
10	rain	mild	normal	false	Y
11	sunny	mild	normal	true	Y
12	overcast	mild	high	true	Y
13	overcast	hot	normal	false	Y
14	rain	mild	high	true	N

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.