Weekly Homework 2

Alex Ring

September 6, 2016

Exercise 10.1 B) Suppose that instead of binary classes (over 60, under 60), your data consists of a discrete set of ages, A, in [0,100] and probabilities pϵ[0,1] for each age, respectively. Write an equation (with variables, not numbers) for the probability of people over age 60.

The probability of someone being over age 60 is the total probability of those ages 60 and over. That is to say:

p(x,c) = p(c)$\prod_{i=60}^{100} p(x_i|c)$

Exercise 10.5
1. Derive expressions for the parameters of this model in terms of the training data using maximum likelihood. Assume that the data is independent and indenticaly distributed.

p(c1,...,cN,x1,...,xN) = $\prod_{n=1}^{N} p(c^n, x^n)$ $\prod_{i=1}^{D} p(xi|C)$

So here we need p(c) = $\frac{number of times class c occurs}{total number of datapoints}$
This data set can be used as described in section 10.3.1 to give us:
Using maximum likelihood p(c—D) where D is the dataset, p(c—D) = $\frac{1}{N}\sum_{n=1}^{N} II[c^n = c]$

2. Given a trained model p(x,c) explain how to form a classifier p(x—c)

According to the reading, in order to form a classifier given a trained model. We are given a trained model:
p(x,c), which can be written as: $p(c|x^*) = \prod_{i=1}^{D} p(x_i|c)$
You would then take this and use Bayes' rule to form the classifier:
$p(c|x^*) = \frac{p(x^*|c)p(c)}{p(x^*)} = \frac{p(x^*|c)p(c)}{\sum_c p(x^*|c)p(c)}$
So the classifier is:
$p(c,x) = \frac{p(x^*|c)p(c)}{\sum_c p(x^*|c)p(c)}$

3. If ’viagra’ never appears in the spam training data, discuss what effect this will have on the classification for a new email that contains the word ’viagra’. Explain how you might counter this effect. Explain how a spammer might try to fool a naive Bayes spam filter.

If the word ’viagra’ has never shown up before in spam training data, it will have a value of 0, not spam. This would make it so the email would probably get classified as not being spam if it were soley based on ’viagra’ being in the contents of the email. In order to counter this you can train the spam filter on generic emails that are trying to sell or enhance things. You would assume a spam email about viagra would contain other words about buying, cash, enchancement, ect. By training the filter on generic emails trying to sell things, you can flag the viagra email as spam with the word viagra having little effect on the classification at all. Spammers could try to fool a naive bayes algorithm by including extraneous things attached to their spam words. For example, if I were to include the word cash in my email, but threw some random html comments in there. IE: Cas¡!– –¿h the algorithm would have a hard time recognizing this word as one that was already associated with spam. Unless there was some kind of handler for this sort of thing. Alternatively spammers could just discover which of their emails are being filtered and use entirely different language the next time.

Dirchlet Questions

A) They way I understand it with the example about the cup of balls, is that you want to take a Dirchlet distribution to eventually see where the data is going to converge. You can take a random bit of the data and eventually you will have a uniform distribution that is representative of that data. In the ball in the cup example you have some colors of balls and every time you take out one you put back that ball and a ball of the same color. Eventually you are going to reach a convergence that is representative of the randomness of what color ball you are going to choose from the cup. By taking the distribution you eventually will find how random the data is.

B) Rank the following in order of entropy: Dir(1,1,1), Dir(2,2,2), and Dir(0.1,0.1,0.1) According to the definition in the book, the more the distribution is like a uniform distribution, the greater the entropy. So from the least entropy to the greatest entropy, we have:

1. Dir(0.1,0.1,0.1)
2. Dir(2,2,2)
3. Dir(1,1,1)