Week 2 Homework

Dalton Rothenberger

Problem 1:

Part A:

Super Market Data

Young

Cornflakes	**0**	**1**
Frosties	1	1
Sugar Puffs	1	1
Branflakes	0	0

Old

Cornflakes	**1**	**1**	**1**	**0**
Frosties	0	0	1	0
Sugar Puffs	0	0	1	0
Branflakes	0	1	1	1

Novel Customer: x = (0110)

Young Probability Calculations:

$p(young) = 0.3333$

$p(cornflakes = dislike|young) = 0.5$

$p(frosties = like|young) = 1$

$p(sugarpuffs = like|young) = 1$

$p(branflakes = dislike|young) = 1$

Old Probability Calculations:

$p(old) = 0.6667$

$p(cornflakes = dislike|old) = 0.25$

$p(frosties = like|old) = 0.25$

$p(sugarpuffs = like|old) = 0.25$

$p(branflakes = dislike|old) = 0.25$

Probability that the novel customer is younger than 60:

$$p(young|x) = \frac{p(young)p(cornflakes{=}dislike|young)p(frosties{=}like|young)p(sugarpuffs{=}like|young)p(branflakes{=}dislike|young)}{numerator+p(old)p(cornflakes{=}dislike|old)p(frosties{=}like|old)p(sugarpuffs{=}like|old)p(branflakes{=}dislike|old)}$$

$$p(young|x) = \frac{0.3333*0.5*1*1*1}{(0.3333*0.5*1*1*1)+(0.6667*0.25*0.25*0.25*0.25)}$$

$$p(young|x) = \frac{0.0833}{0.0859}$$

$$p(young|x) = 0.9697$$

Part B:

$\sum_{i=60}^{100} p(i)$

Problem 2:

Part A:

1.
   - $p(c = 1) = \frac{count(c=1)}{count(c=1)+count(c=0)}$
   - $p(x_i = 1|c = 1) = \frac{count(x_i=1,c=1)}{count(c=1)}$
   - $p(x_i = 1|c = 0) = \frac{count(x_i=1,c=0)}{count(c=0)}$
2. Using Naïve Bayes model you determine whether the probability of spam is greater than the probability of not spam. If the probability is greater for SPAM is greater than NOT_SPAM, select SPAM. If the probability for SPAM is less than NOT_SPAM, select NOT_SPAM.
3. If the word 'Viagra' has never been seen then the probability of it occurring is 0 and since the Naïve Bayes formula uses multiplication, anything multiplied by 0 is 0. To counter this, you would use smoothing. To do this you would add a flat value to all the probabilities so that you never end up with a case of multiplying by 0. To fool a spam filter, a spammer could put gibberish at the end of the email to cause the case of multiplying by 0 in systems without smoothing.

Problem 3:

Part A:

If we setup a Dirichlet distribution such that α is the values of the priors for then the values inside the distributions would be the posteriors. This would be useful for cases where we have more than 2 possible classifications because they all have their own distributions inside the Dirichlet distribution with the prior as their value in α for the Dirichlet distribution. This means when we sample the Dirichlet distribution, the probabilities for each classification is imitated by α and once a distribution is sampled we have access to the probabilities that are inside of it.

Part B:

Highest Entropy	$Dir(< 1, 1, 1 >)$
	$Dir(< 2, 2, 2 >)$
Lowest Entropy	$Dir(< 0.1, 0.1, 0.1 >)$

The ‘1’ Dir has the most entropy because it is most uniformly distributed of them all. The ‘2’ Dir is in the middle because as the values in the Dir increase over 1, the density becomes symmetric around the center of the simplex. This means the Dir with ‘2’ has its density around the center. The ‘0.1’ Dir has the least entropy because it has sharp peaks of density at the vertices of the simplex and low density everywhere else making it compact.