

# CS 4780/5780 Homework 4

Due: Sunday 10/18/18 11:55pm on Gradescope

## Problem 1: Intuition for Naive Bayes

Ronnie is playing a game named Flippin' Extravaganza, he asks if someone can help him win. Can you help him beat the house?

a)

This game is easy to undertake, there is a red hat and a blue hat with a weighted penny respectively. The operator secretly flips one of them and asks you to guess under which hat it is. Suppose you know that the red hat's penny is weighted to come up heads  $3/5$  of the time, and the blue hat's penny is weighted to come up heads  $7/10$  of the time. If the penny comes up heads, what is the probability that it came from the red hat?

b)

To get you hooked, the operator next makes the game more complex, but with better odds. Each hat actually has a penny, a nickel, a dime, and a quarter – all weighted. The operator secretly selects all the coins from one hat at random, flips them all, and asks you to guess which hat they came from. Suppose you know the the red hat's penny, nickel, dime, and quarter come up heads with probability  $[3/5, 3/10, 1/2, 4/5]$ , respectively, and the blue hat's coins have heads probabilities  $[7/10, 1/5, 1/10, 2/5]$ , respectively. If the coins come up  $[H, H, T, H]$ , what is the probability that they came from the red hat?

c)

Previously we made the assumption that you somehow know the coins' weights. In fact, in the Coin Flippin' Extravaganza, the operator never reveals the weights. Instead, you've spent the whole day watching people play the game, recording the results below:

| game | penny | nickel | dime | quarter | hat  |
|------|-------|--------|------|---------|------|
| 1    | T     | H      | T    | T       | Red  |
| 2    | T     | T      | H    | T       | Blue |
| 3    | T     | H      | T    | H       | Blue |
| 4    | H     | H      | H    | T       | Red  |
| 5    | H     | H      | T    | T       | Red  |
| 6    | T     | T      | H    | H       | Blue |
| 7    | H     | H      | T    | T       | Red  |
| 8    | T     | T      | H    | T       | Blue |
| 9    | T     | H      | H    | T       | Blue |
| 10   | H     | H      | H    | T       | Red  |
| 11   | T     | T      | H    | T       | Blue |
| 12   | T     | H      | H    | T       | Red  |
| 13   | H     | H      | T    | T       | Red  |
| 14   | T     | T      | H    | H       | Blue |
| 15   | T     | H      | H    | T       | Blue |
| 16   | T     | T      | H    | H       | Blue |
| 17   | H     | T      | H    | H       | Red  |
| 18   | H     | T      | H    | T       | Blue |

i. Estimate the following probabilities with +1 smoothing

| hat  | $P(\text{penny} = H \text{hat})$ | $P(\text{nickel} = H \text{hat})$ | $P(\text{dime} = H \text{hat})$ | $P(\text{quarter} = H \text{hat})$ |
|------|----------------------------------|-----------------------------------|---------------------------------|------------------------------------|
| Red  |                                  |                                   |                                 |                                    |
| Blue |                                  |                                   |                                 |                                    |

ii. Now if the coins come up  $[H, H, T, H]$ , what is the probability that they came from the red hat?

d)

The above problem is actually a Naive Bayes problem with Bernoulli distributed features. To see this, define the feature space  $\mathcal{X}$  and the label space  $\mathcal{Y}$ . Is the Naive Bayes assumption valid for this problem? Why?

## Problem 2: Linearity of Gaussian Naive Bayes

In this question, you will show that Naive Bayes is a linear classifier when using Gaussian likelihood factors with shared variances. Specifically, consider the following Naive Bayes model:

$$p(y|\mathbf{x}) = \frac{\prod_{\alpha=1}^d p([\mathbf{x}]_{\alpha}|y) p(y)}{p(x)}$$

with:

$$p([\mathbf{x}]_{\alpha}|y) = \mathcal{N}([\mu_y]_{\alpha}, [\sigma]_{\alpha})$$

That is, there is a separate mean value for each feature  $\mathbf{x}_{\alpha}$  and each class  $y \in \{0, 1\}$ . However, variances are shared across classes, so that there is only one variance  $[\sigma]_{\alpha}$  per feature.

a)

Show that the decision rule  $p(y = 1|x)$  can equivalently be written as:

$$p(y = 1|\mathbf{x}) = \frac{\prod_{\alpha=1}^d p([\mathbf{x}]_{\alpha}|y = 1)p(y = 1)}{\prod_{\alpha=1}^d p([\mathbf{x}]_{\alpha}|y = 1)p(y = 1) + \prod_{\alpha=1}^d p([\mathbf{x}]_{\alpha}|y = 0)p(y = 0)}$$

Hint: remember the sum rule and the product rule.

b)

Using this formulation, show how to rewrite  $p(y = 1|x)$  as:

$$p(y = 1|\mathbf{x}) = \frac{1}{1 + \exp\left(-\log \frac{\prod_{\alpha=1}^d p([\mathbf{x}]_{\alpha}|y=1)p(y=1)}{\prod_{\alpha=1}^d p([\mathbf{x}]_{\alpha}|y=0)p(y=0)}\right)}$$

c)

Given the above expression for  $p(y = 1|x)$ , show that Naive Bayes with this definition of  $p([\mathbf{x}]_{\alpha}|y)$  is a linear model. Hint: the form you derived in part b should remind you of a decision rule you have seen before.