

CS 4780/5780 Homework 4

Due: Tuesday 03/06/18 11:55pm on Gradescope

Problem 1: Intuition for naive Bayes

Kilian loves carnivals and brings the whole class along for his yearly trip. His favorite game is the Coin Flippin' Extravaganza, and he proclaims that any student that can help him win will get an A in his heart. Can you help him beat the house?

a)

The game starts simple. There is a red hat and a blue hat each with a weighted penny. The operator secretly selects a penny at random, flips it, and asks you to guess which hat it came from. 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 easier 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)

We've been making 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

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: Usefulness of naive Bayes

We wish to learn a classifier from the following dataset, in order to sort future emails as spam emails or as emails pertaining to ham. The dataset contains 15 emails, 3 binary features for each email, and the binary label.

id	has word "bacon"	from recognized IP	contains misspelled word	label
1	0	0	1	spam
2	0	0	1	spam
3	0	1	0	spam
4	0	0	1	spam
5	1	0	0	spam
6	0	1	0	spam
7	0	0	1	spam
8	0	1	1	spam
9	0	0	1	spam
10	0	0	1	spam
11	1	1	0	ham
12	1	1	1	ham
13	1	0	1	ham
14	1	0	1	ham
15	1	1	0	ham

a)

Let's start with a simple model, where for each combination of features there is some underlying probability that the email pertains to ham. Using maximum likelihood estimation, what is the probability that an email pertains to ham if its feature vector is $(0, 0, 1)$? What if the feature vector is $(1, 1, 1)$? Or $(1, 0, 0)$? How about $(0, 0, 0)$? Are any of the predictions undefined? Do any of them seem unreasonable? If so, why?

b)

Consider the following ideas for improving our model. For each one, explain why you think the idea would or would not help to make our predictions well-defined or more reasonable.

- Collect more emails for the training dataset.
- Extract more features for each email.
- Duplicate emails with uncommon combinations of features.
- Make stronger assumptions in our generative model for the data.

c)

Let's try the last idea and make stronger model assumptions. Specifically, let's make the naive Bayes assumption that feature values are independent given the labels. Model each conditional probability with a Bernoulli distribution. What is the probability that an email pertaining to ham has the feature vector $(1, 0, 1)$? (Hint: start by using MLE to estimate the probability that an email has the word "bacon" given that it pertains to ham.)

d)

Using the naive Bayes classifier with the distributions you found above, what is the probability that an email pertains to ham if its feature vector is $(0, 0, 1)$? What if the feature vector is $(1, 1, 1)$? Or $(1, 0, 0)$? How about $(0, 0, 0)$? Are any of the predictions undefined? Do any of them seem unreasonable? If so, why?

e)

Complete parts **c** and **d** above, this time using +1 additive smoothing when estimating the model parameters. (Fun fact: this is effectively doing MAP estimation with a Beta prior distribution.) Do any of the predictions change? Do the probabilities?

Problem 3: 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:

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

with:

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

That is, there is a separate mean value for each feature $[x]_{\alpha}$ 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:

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

Hint: remember the sum rule and the product rule.

b)

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

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

c)

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