CS 188
Summer 2019

Section Handout 12

Naive Bayes

$$prediction(x_1, ..., x_n) = \arg\max_y P(Y = y) \prod_{i=1}^{n} P(X_i = x_i | Y = y)$$

Parameter Estimation

Given sample your **maximum likelihood** estimate for an outcome x that can take on $|X|$ different values from a sample of size N is

$$P_{MLE}(x) = \frac{count(x)}{N}.$$

With **Laplace smoothing,** the Laplace estimate with strength k is

$$P_{LAP,k}(x) = \frac{count(x) + k}{N + k|X|}.$$

A similar result holds for computing Laplace estimates for conditionals (which is useful for computing Laplace estimates for outcomes across different classes):

$$P_{LAP,k}(x|y) = \frac{count(x, y) + k}{count(y) + k|X|}.$$

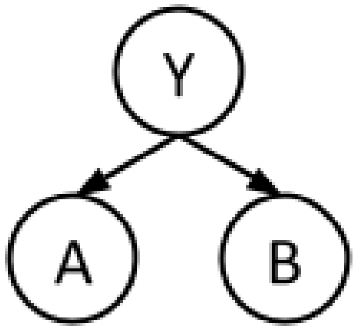
There are two particularly notable cases for Laplace smoothing. The first is when $k = 0$, then $P_{LAP,0}(x) = P_{MLE}(x)$. The second is the case where $k = \infty$. Observing a very large, infinite number of each outcome makes the results of your actual sample inconsequential and so your Laplace estimates imply that each outcome is equally likely. Indeed:

$$P_{LAP,\infty}(x) = \frac{1}{|X|}$$

1 Naive Bayes

In this question, we will train a Naive Bayes classifier to predict class labels Y as a function of input features A and B. Y, A, and B are all binary variables, with domains 0 and 1. We are given 10 training points from which we will estimate our distribution.

A	1	1	1	1	0	1	0	1	1	1
B	1	0	0	1	1	1	1	0	1	1
Y	1	1	0	0	0	1	1	0	0	0



(a) What are the maximum likelihood estimates for the tables $P(Y)$, $P(A|Y)$, and $P(B|Y)$?

Y	$P(Y)$
0	
1	

A	Y	$P(A\|Y)$
0	0	
1	0	
0	1	
1	1	

B	Y	$P(B\|Y)$
0	0	
1	0	
0	1	
1	1	

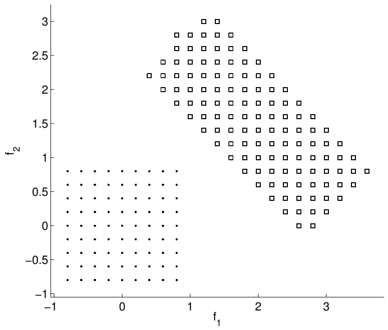
(b) Consider a new data point ($A = 1$, $B = 1$). What label would this classifier assign to this sample?

(c) Let's use Laplace Smoothing to smooth out our distribution. Compute the new distribution for $P(A|Y)$ given Laplace Smoothing with $k = 2$.

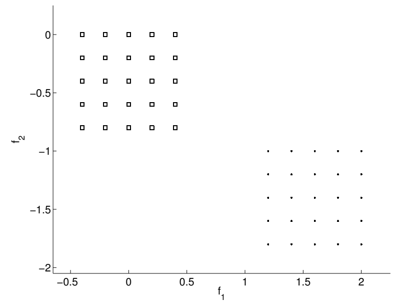
A	Y	$P(A\|Y)$
0	0	
1	0	
0	1	
1	1	

Q2. Naïve Bayes Modeling Assumptions

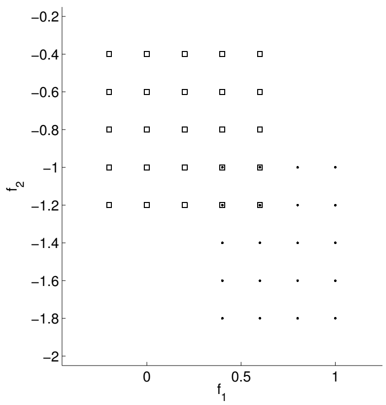
You are given points from 2 classes, shown as rectangles and dots. For each of the following sets of points, mark if they satisfy all the Naïve Bayes modelling assumptions, or they do not satisfy all the Naïve Bayes modelling assumptions. Note that in (c), 4 rectangles overlap with 4 dots.



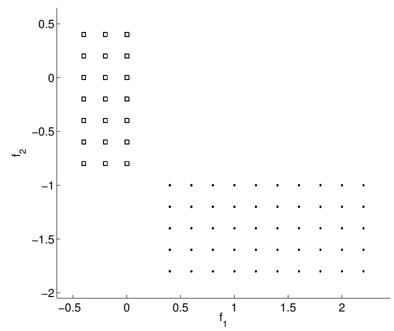
(a) Satisfies Does not Satisfy



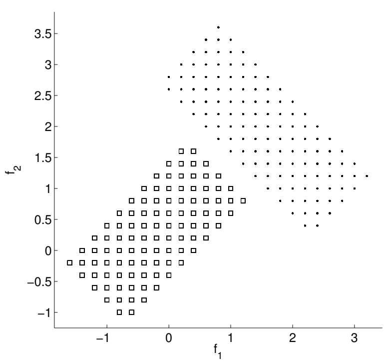
(b) Satisfies Does not Satisfy



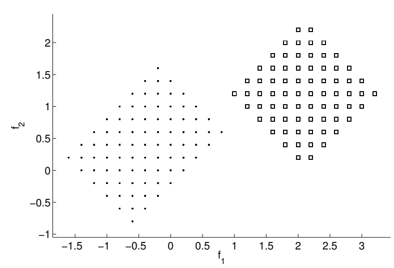
(c) Satisfies Does not Satisfy



(d) Satisfies Does not Satisfy



(e) Satisfies Does not Satisfy



(f) Satisfies Does not Satisfy

3 Maximum Likelihood

A Geometric distribution is a probability distribution of the number X of Bernoulli trials needed to get one success. It depends on a parameter p, which is the probability of success for each individual Bernoulli trial. Think of it as the number of times you must flip a coin before flipping heads. The probability is given as follows:

$$P(X = k) = p(1 - p)^{k-1} \tag{1}$$

p is the parameter we wish to estimate.

We observe the following samples from a Geometric distribution: $x_1 = 5$, $x_2 = 8$, $x_3 = 3$, $x_4 = 5$, $x_5 = 7$. What is the maximum likelihood estimate for p?

Q4. Generalization

(a) Suppose you train a classifier and test it on a held-out validation set. It gets 80% classification accuracy on the training set and 20% classification accuracy on the validation set.

From what problem is your model most likely suffering?

◯ Underfitting ◯ Overfitting

Fill in the bubble next to any measure of the following which could reasonably be expected to improve your classifier's performance on the validation set.

◯ Add extra features ◯ Remove some features

Briefly justify:

◯ Collect more training data ◯ Throw out some training data

Assuming features are outcome counts (k is the Laplace smoothing parameter controlling the number of extra times you "pretend" to have seen an outcome in the training data):

◯ Increase k ◯ Decrease k (assuming $k > 0$ currently)

Assuming your classifier is a Bayes' net:

◯ Add edges ◯ Remove edges

(b) Suppose you train a classifier and test it on a held-out validation set. It gets 30% classification accuracy on the training set and 30% classification accuracy on the validation set.

From what problem is your model most likely suffering?

◯ Underfitting ◯ Overfitting

Fill in the bubble next to any measure of the following which could reasonably be expected to improve your classifier's performance on the validation set.

◯ Add extra features ◯ Remove some features

Briefly justify:

◯ Collect more training data ◯ Throw out some training data

(c) Your boss provides you with an image dataset in which some of the images contain your company's logo, and others contain competitors' logos. You are tasked to code up a classifier to distinguish your company's logos from competitors' logos. You complete the assignment quickly and even send your boss your code for training the classifier, but your boss is furious. Your boss says that when running your code with images and a random label for each of the images as input, the classifier achieved perfect accuracy on the training set. And this happens for all of the many random labelings that were generated.

Do you agree that this is a problem? Justify your answer.