CS 445/545, Machine Learning, Spring, 2013

Homework 4: Naïve Bayes Classification

Due Monday June 3, 2:00pm

I. Written part

1. Consider the following training set, in which each example has four binary attributes and a binary class.

Example	Attribute1	Attribute2	Attribute3	Attribute4	Class
$\mathbf{x}_1$	1	1	1	1	+1
$\mathbf{x}_2$	1	1	0	1	+1
$\mathbf{x}_3$	0	1	1	0	+1
$\mathbf{x}_4$	1	0	0	1	+1
$\mathbf{x}_5$	1	0	0	0	−1
$\mathbf{x}_6$	1	0	1	0	−1
$\mathbf{x}_7$	0	1	0	0	−1
$\mathbf{x}_8$	0	0	1	0	−1

(a) For each attribute, each value, and each class, give the smoothed probability, P(Attribute=value | class), using Laplace (add-one) smoothing, as described in class.

(b) Show how a naïve Bayes classifier trained on this training set would classify the following new example:

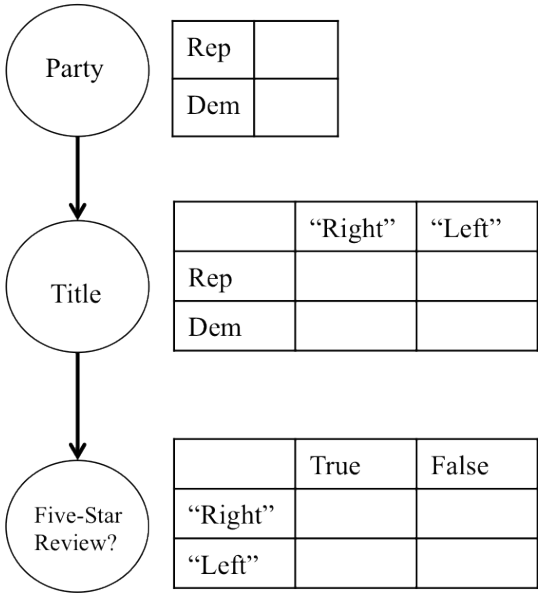
$\mathbf{x}_9$	1	1	0	0	?

(Show your work.)

2. You have landed a Data Scientist job at Amazon.com and have been doing a study on Republican and Democratic political authors and their books' titles. In particular, you have been looking at a collection of books whose titles contain the terms "Right" and "Left". 70% of these books' authors are Republican and 30% are Democrats. You have found that if a book's author is a Republican, there is a 90% probability that the title contains "Left " and a 10% probability that the title contains "Right." Amazingly, you have also found that if the book's author is a Democrat, then exactly the opposite is true: there is a 90% probability that the title contains "Right" and a 10% probability that the title contains "Left". (You haven't yet seen any books with both terms in the title.)

You have also found that titles with the term "Right" have an 80% probability of at least one 5-star review, whereas titles with the term "Left" have only a 50% chance of at least one 5-star review.

(a) Given the Bayesian Network shown below, fill in the conditional probability tables to reflect the information above.



(b) Suppose a book in your study has "Left" in the title. What is the probability that it was written by a Democrat? Show your work.

(c) Suppose a book in your study has a five-star review and "Right" in the title. What is the probability that it was written by a Democrat? Show your work.

II. Computing part

In this part of the homework you will experiment on using Naïve Bayes to classify the Optdigits data, and compare it to your results from Homework 1 (Decision Trees).

Dataset: The OptDigits dataset from Homework 1. Recall that this has 64 attributes, each of which can have value 0−16.

1. Implement Naïve Bayes and learn model from the training data:

Use the Naïve Bayes algorithm to create a model from the training set (optdigits.train) as described in class:

A. Compute the prior probability for each class. For example P('1') is the probability of digit '1' in the training set.

B. Compute conditional probabilities for each digit and for each value of each attribute. For example, P(Attribute1=0 | '1') is the probability of that Attribute 1 has value 0, given that the class is digit '1'.

C. Smooth the conditional probabilities using Laplace (add-one) smoothing, as described in class.

2. Run Naïve Bayes on the test data:

- Use the Naïve Bayes algorithm to classify the instances in optdigits.test. Because a product of 65 probabilities will be very small, we will instead use the log of the product. Recall that the classification method is:

$$class_{NB}(\mathbf{x}) = \underset{class}{\operatorname{argmax}} \left[P(class) \prod_i P(x_i \mid class) \right]$$

Since

$$\underset{z}{\operatorname{argmax}}\, f(z) = \underset{z}{\operatorname{argmax}} \log f(z)$$

we have:

$$class_{NB}(\mathbf{x}) = \underset{class}{\operatorname{argmax}} \log \left[P(class) \prod_i P(x_i \mid class) \right]$$

$$= \underset{class}{\operatorname{argmax}} \left[\log P(class) + \log P(x_1 \mid class) + \ldots + \log P(x_n \mid class) \right]$$

Give the accuracy on the test set, and a single confusion matrix for all 10 digits. Write a few sentences describing your results, and comparing your results here with the results you obtained using decision trees on this problem (in Homework 1). Do you think the attributes here are independent, as assumed by Naïve Bayes? Does Naïve Bayes do well on this problem in spite of the independence assumption? Speculate on other reasons Naïve Bayes might do well or poorly on this problem.

3. Optional: Redo Step 1 **B** and **C** using binning of the attributes into 4 discrete values instead of 17 – that is, for value v of each attribute, there are 4 bins:

$0 <= v <= 4$
$4 < v <= 8$
$8 < v <= 12$
$12 < v <= 16$

Use these bins to redo step 2. Does this coarser binning improve your results?

What to turn in:
Your answers to the written part and your results and discussion of the computation part.

How to turn it in:

Part I (written) can be turned in either in hard copy or electronically. Part II (computing): Please turn in an electronic version of your results (accuracy and confusion matrix) and discussion, as well as an electronic version of your Naïve Bayes code. Send these to mm@cs.pdx.edu.

If there are any questions on this assignment, don't hesitate to ask me, Max (our TA), or e-mail the class mailing list.

Policy on late homework: If you are having trouble completing the assignment on time for any reason, please see me before the due date to find out if you can get an extension. Any homework turned in late without an extension from me will have 5% of the grade subtracted for each day the assignment is late.