

Machine Learning

Assignment 4

Daniel Leblanc

June 3, 2013

Part I:

1: (a) Performing the Laplace smoothing on binary attributes means just adding 1 to the numerators and 2 to the denominators. This results in the following smoothed probabilities:

$$\begin{array}{ll} Pr(A1 = 0|+) = \frac{2}{6} & Pr(A1 = 1|+) = \frac{4}{6} \\ Pr(A2 = 0|+) = \frac{2}{6} & Pr(A2 = 1|+) = \frac{4}{6} \\ Pr(A3 = 0|+) = \frac{3}{6} & Pr(A3 = 1|+) = \frac{3}{6} \\ Pr(A4 = 0|+) = \frac{2}{6} & Pr(A4 = 1|+) = \frac{4}{6} \\ Pr(A1 = 0|-) = \frac{3}{6} & Pr(A1 = 1|-) = \frac{3}{6} \\ Pr(A2 = 0|-) = \frac{4}{6} & Pr(A2 = 1|-) = \frac{2}{6} \\ Pr(A3 = 0|-) = \frac{3}{6} & Pr(A3 = 1|-) = \frac{3}{6} \\ Pr(A4 = 0|-) = \frac{5}{6} & Pr(A4 = 1|-) = \frac{1}{6} \end{array}$$

1: (b) For the +1 class:

$$\frac{4}{6} \cdot \frac{4}{6} \cdot \frac{3}{6} \cdot \frac{2}{6} = \frac{96}{1296} \approx 0.0741$$

For the -1 class:

$$\frac{3}{6} \cdot \frac{2}{6} \cdot \frac{3}{6} \cdot \frac{5}{6} = \frac{90}{1296} \approx 0.0694$$

So x_9 would be classified as +1.

2: (a) For this question I just needed to fill in the values that are provided in the question:

Rep	.7
Dem	.3

	“Right”	“Left”
Rep	.1	.9
Dem	.9	.1

	True	False
“Right”	.8	.2
“Left”	.5	.5

2: (b) Using condition probability I get the following:

$$Pr(Dem|“Left”) = \frac{Pr(“Left”|Dem)Pr(Dem)}{Pr(“Left”)}$$

Since we don’t have the $Pr(“Left”)$ we’ll need to calculate that. It is:

$$Pr(“Left”) = Pr(“Left”|Rep)Pr(Rep) + Pr(“Left”|Dem)Pr(Dem) = .9 \cdot .7 + .1 \cdot .3 = 0.66$$

Plugging in the values from the previous step I get:

$$Pr(Dem|“Left”) = \frac{.1 \cdot .3}{.66} = .045455$$

2: (c) The important piece to note here, is that the probability of a five-star review is independent of whether it was written by a Republican or a Democrat. So for our purposes we can remove that information and save it until the end.

$$Pr(Dem|“Right”) = \frac{Pr(“Right”|Dem)Pr(Dem)}{Pr(“Right”)} = \frac{.9 \cdot .3}{.34} = .794118$$

To get our final result we just need to multiply that by the probability that there is a five-star review given that the title contains “Right”.

$$Pr(Dem|“Right”)Pr(True|“Right”) = .794118 \cdot .8 = .625294$$

Part II:

All code for question 1 and 2 are in the file *bayes.py*. The resulting accuracy on the test set was:

$$Accuracy = 90.04\%$$

The confusion Matrix was:

	0	1	2	3	4	5	6	7	8	9
0	91	0	0	0	2	0	0	0	0	2
1	0	82	0	0	0	0	1	1	6	1
2	0	1	86	0	0	0	0	0	5	2
3	0	0	0	97	0	1	0	0	1	6
4	0	1	0	0	70	0	0	8	1	4
5	0	1	1	1	1	73	1	0	3	7
6	0	2	0	0	0	0	89	0	0	0
7	0	0	1	1	0	0	0	89	3	0
8	0	2	0	5	0	3	0	0	76	2
9	0	0	0	6	3	0	0	5	2	88

Overall the results were much better than I predicted considering how simple the Naive Bayesian classifier is. The final accuracy was about 10% better than the best results I got from the decision trees in assignment 1. The confusions that did occur were similar to the ones that resulted from the decision tree, just much less frequent. The attributes are certainly not independent. Each attribute value depends greatly on the neighboring values just by virtue of pen motions. This realistically doesn't much matter, the lack of independence is balanced across all 10 digits. Even though dependence errors are introduced, those errors are introduced for all of the digits, resulting in a classifier that still performs well.