DSC 40A Fall 2025 - Group Work Week 10

due Monday, Dec 1st at 11:59PM

Write your solutions to the following problems by either typing them up or handwriting them on another piece of paper. **One person** from each group should submit your solutions to Gradescope and **tag all group members** so everyone gets credit.

This worksheet won't be graded on correctness, but rather on good-faith effort. Even if you don't solve any of the problems, you should include some explanation of what you thought about and discussed, so that you can get credit for spending time on the assignment.

In order to receive full credit, you must work in a group of two to four students for at least 50 minutes in your assigned discussion section. You can also self-organize a group and meet outside of discussion section for 80 percent credit. You may not do the groupwork alone.

Problem 1.

The table below contains 12 short movie reviews that have been manually labeled as either positive ($y = 1$) or negative ($y = 0$). For every review five binary features are recorded indicating whether the corresponding word appears in the text:

review i	`great`	`boring`	`plot`	`acting`	`slow`	y
1	1	0	1	1	0	1
2	1	0	1	0	0	1
3	0	0	1	1	0	1
4	1	0	0	1	0	1
5	0	1	0	0	1	0
6	0	1	1	0	1	0
7	0	0	0	1	1	0
8	1	0	0	0	1	0
9	0	1	1	1	0	0
10	1	1	1	0	0	0
11	0	1	1	1	0	1
12	0	0	1	0	1	1

Let the feature vector $\vec{X} = (X^{(1)}, X^{(2)}, X^{(3)}, X^{(4)}, X^{(5)})$ correspond to the presence of the words `great`, `boring`, `plot`, `acting`, `slow`, respectively, and let $Y \in \{0, 1\}$ denote the sentiment class.

a) Compute the probabilities $\Pr(\{Y = 1\})$ and $\Pr(\{Y = 0\})$ from the table.

b) Consider a new review whose feature vector is

$$\vec{x} = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \end{bmatrix}^\top .$$

Using the naïve Bayes classifier, compute the class likelihoods $\Pr\left(\{Y = 0\} \mid \{\vec{X} = \vec{x}\}\right)$ and $\Pr\left(\{Y = 1\} \mid \{\vec{X} = \vec{x}\}\right)$. If you need to use Laplace smoothing, clearly indicate why.

c) Which class does the classifier predict for this review?

Problem 2.

Researchers on planet ZOG are building a classifier that decides whether a brand new jelly bean is YUMMY ($y = 1$) or YUCKY ($y = 0$). For each candy they record five quirky binary features:

Feature j	Meaning of $\{X^{(j)} = 1\}$
1	the bean glows in the dark (`glow`)
2	the bean fizzes when bitten (`fizz`)
3	the bean feels slimy (`slimy`)
4	the bean crunches loudly (`crunch`)
5	the bean whistles when shaken (`whistle`)

The training data of 14 beans are listed below.

bean i	glow	fizz	slimy	crunch	whistle	y
1	1	1	0	1	0	1
2	1	0	0	1	0	1
3	0	1	0	0	0	1
4	1	1	0	0	0	1
5	0	0	1	0	1	0
6	0	1	1	1	0	0
7	0	0	1	0	0	0
8	1	0	1	0	1	0
9	0	1	1	0	1	0
10	1	0	1	1	1	0
11	0	1	0	1	0	1
12	1	0	0	0	0	1
13	0	0	1	1	1	0
14	1	1	1	0	1	0

Let $\vec{X} = (X^{(1)}, \ldots, X^{(5)})$ denote the five features and $Y \in \{0, 1\}$ the label.

a) Compute the probabilities $\Pr(\{Y = 1\})$ and $\Pr(\{Y = 0\})$ from the table.

b) A brand-new jelly-bean has feature vector

$$\vec{x} = \begin{bmatrix} 0 & 1 & 0 & 0 & 1 \end{bmatrix}^\top \quad (\texttt{fizz} + \texttt{whistle} \text{ only}).$$

Estimate the probabilities $\Pr\left(\{Y = 0\} \mid \{\vec{X} = \vec{x}\}\right)$ and $\Pr\left(\{Y = 1\} \mid \{\vec{X} = \vec{x}\}\right)$. If you need to use Laplace smoothing, clearly indicate why.

c) Which class does the classifier predict for this jelly bean?

Problem 3. Naïve Bayes as a linear classifier

Consider a binary classification problem with label $Y \in \{0, 1\}$ and d binary features $\vec{X} = (X^{(1)}, \ldots, X^{(d)})$, where each $X^{(j)} \in \{0, 1\}$. Assume the naïve Bayes model with $\{0, 1\}$ features:

$$\Pr(\{Y = y\}) = \pi_y, \quad \Pr\left(\{X^{(j)} = 1 \mid Y = y\}\right) = \theta_{j,y}, \quad j = 1, \ldots, d, \ y \in \{0, 1\},$$

and that, conditional on Y, the features $X^{(1)}, \ldots, X^{(d)}$ are independent.

a) Show that the "log-posterior odds" can be written as

$$\log \frac{\Pr\left(\{Y = 1 \mid \vec{X} = \vec{x}\}\right)}{\Pr\left(\{Y = 0 \mid \vec{X} = \vec{x}\}\right)} = w_0 + \sum_{j=1}^{d} w_j x^{(j)},$$

for suitable constants $w_0, w_1, \ldots, w_d$ that depend only on (π_y) and $(\theta_{j,y})$. Give explicit formulas for w_0 and w_j.

b) Conclude that the naïve Bayes classifier with Bernoulli features has a *linear* decision rule of the form

$$\hat{y}(\vec{x}) = \begin{cases} 1 & \text{if } w_0 + \sum_{j=1}^{d} w_j x^{(j)} \geq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Explain in words what this means geometrically about the decision boundary in $\{0,1\}^d$.

c) Suppose for some feature j we have $\theta_{j,1} = \theta_{j,0}$. Show that $w_j = 0$ in this case and interpret this fact: what does naïve Bayes do with a feature that is *equally distributed* in both classes?