DSC 40A - Homework 06

Due: Friday, February 21, 2020

Write your solutions to the following problems by either typing them up or handwriting them on another piece of paper. Unless otherwise noted by the problem's instructions, show your work or provide some justification for your answer. Homeworks are due via Gradescope on Friday afternoon at 5:00 p.m.

Problem 1.

Let $\vec{x}^{(1)}, \ldots, \vec{x}^{(n)}$ be a set of n data points with associated labels $y_1, \ldots, y_n \in \{-1, 1\}$. Recall from lecture that the perceptron update rule is

$$\vec{w}^{(t)} = \vec{w}^{(t-1)} - \frac{\alpha}{n} \sum_{i \in M} \begin{cases} \text{Aug}(\vec{x}^{(i)}), & \text{Aug}(\vec{x}^{(i)}) \cdot \vec{w}^{(t-1)} \geq 0 \\ -\text{Aug}(\vec{x}^{(i)}), & \text{Aug}(\vec{x}^{(i)}) \cdot \vec{w}^{(t-1)} < 0 \end{cases}$$

where M is the set of data points which are currently misclassified by $\vec{w}^{(t-1)}$.

Suppose we gather the following data:

$$\begin{array}{ll} x_1 = (1,0)^\intercal & y_1 = 1 \\ x_2 = (0,1)^\intercal & y_2 = 1 \\ x_3 = (-1,1)^\intercal & y_3 = 1 \\ x_4 = (1,2)^\intercal & y_4 = -1 \\ x_5 = (2,2)^\intercal & y_5 = -1 \\ x_6 = (2,1)^\intercal & y_6 = -1 \end{array}$$

Run the perceptron algorithm by hand with w initialized to $(1,1,1)^\intercal$, and with a step size of $\alpha = 3$. Stop the algorithm when all points are classified correctly. What is the weight vector w at each step? How many training examples are misclassified at each step? For the purpose of this problem, a training example which lies on the decision boundary is considered to be misclassified.

Solution: Step 0
Number of misclassified points: 3
Current weight vector: [1 1 1]

Step 1
Number of misclassified points: 3
Current weight vector: [-0.5 -1.5 -1.5]

Step 2
Number of misclassified points: 1
Current weight vector: [1. -1.5 -0.5]

Step 3
Number of misclassified points: 0
Current weight vector: [1.5 -1. -0.5]

Problem 2.

Zelda's nightmare has come true: she thought she had dropped DSC 40A before the quarter started, but she is somehow still enrolled. She now has to take the midterm without knowing any of the material. The first section of the midterm consists of four True/False questions. Zelda will guess on each question by picking True/False with equal probability.

Showing your work, compute the probability that

a) Zelda answers True to all of the questions.

Solution: The sample space consists of all possible sequences of responses. For instance, it contains $(T,T,T,T),(T,T,T,F),(F,T,F,F)$, etc. There are $4 \cdot 4 \cdot 4 \cdot 4 = 1/16$ possible outcomes in the sample space, and only one of them is (T,T,T,T). Hence the probability is $\frac{1}{16}$.

b) Zelda answers True to at least one of the questions.

Solution: This is the complement of the event that she answers False to all of the questions. The probability of answering False four times in a row is $\frac{1}{2^4} = \frac{1}{16}$. Hence the probability of answering True at least once is $1 - \frac{1}{16} = \frac{15}{16}$.

c) Zelda gets all of the questions right.

Solution: The sample space consists of all possible sequences of marks: for instance (correct, correct, incorrect, correct), (correct, correct, correct, correct), etc.

There are 16 outcomes in the sample space, and only one of them is (correct, correct, correct, correct). Hence the probability of getting all four right is also $\frac{1}{2^4} = \frac{1}{16}$.

d) Zelda gets none of the questions right.

Solution: Same as the above: There are 16 outcomes in the sample space, and only one of them is (incorrect, incorrect, incorrect, incorrect). Hence the probability of getting all four right is also $\frac{1}{2^4} = \frac{1}{16}$.

e) Zelda gets at least one question right.

Solution: This is the complement of getting all of the questions wrong. The probability of that is $\frac{1}{16}$. Hence the probability of getting at least one right is $1 - \frac{1}{16} = \frac{15}{16}$.

f) Zelda gets exactly one of the questions right.

Solution: There are four outcomes in this event: (C,I,I,I), (I,C,I,I), (I,I,C,I), and (I,I,I,C). There are 16 equiprobable outcomes, hence the probability is $4/16 = 1/4$.

g) Zelda gets exactly half of the questions right.

Solution: There are six outcomes in which Zelda gets half of the questions right: (C,C,I,I), (C,I,C,I), (C,I,I,C), (I,C,C,I), (I,C,I,C), (I,I,C,C). Hence the probability is $6/16 = 3/8$.

h) the first question Zelda gets right is Question 3.

Solution: There are two outcomes where the third question is the first one that Zelda gets right: (I,I,C,C), (I,I,C,I). Hence the probability is $2/16 = 1/8$.

Problem 3.

Now Zelda has moved on to the next question which consists of four multiple choice questions. Each question has a different number of choices:

Question	Number of Choices
1	3
2	4
3	2
4	5

Zelda will guess randomly on each question, as before. Each guess is independent of the one before it.

Showing your work, compute the probability that

a) Zelda gets all of the questions right.

Solution: The sample space consists of sequences of the form:

(first answer, second answer, third answer, fourth answer)

This is the Cartesian product of four sets, each set being the possible choices for that question. The total size of the sample space is therefore $3 \cdot 4 \cdot 2 \cdot 5 = 120$.

There is only one outcome in which Zelda gets all four questions right, and so the probability of this event is $\frac{1}{120}$.

b) Zelda gets none of the questions right.

Solution: The sample space has 120 entries. There are 2 wrong answers to the first question, 3 wrong answers to the second, 1 wrong answer to the third, and 4 wrong answers to the fifth. The set of outcomes in which Zelda gets none of the questions right is the Cartesian product of the sets of wrong answers to each question. This product will have $2 \cdot 3 \cdot 1 \cdot 4 = 24$ elements. Hence the probability is $24/120$.

c) Zelda gets at least one question right.

Solution: This is the complement of getting none of the questions right.

The probability of getting no questions right is

$$\frac{2}{3} \cdot \frac{3}{4} \cdot \frac{1}{2} \cdot \frac{4}{5} = \frac{24}{120} = \frac{3}{15}$$

So the probability of getting at least one question right is:

$$1 - \frac{3}{15} = \frac{12}{15}$$

d) Zelda gets exactly one of the questions right.

Solution: This event consists of the four outcomes: (C,I,I,I), (I,C,I,I), (I,I,C,I), and (I,I,I,C).

Each outcome is mutually exclusive, so the probability of the event is the sum of the probabilities of each outcome.

The probability of (C,I,I,I) is: $\frac{1}{3} \cdot \frac{3}{4} \cdot \frac{1}{2} \cdot \frac{4}{5} = \frac{12}{120}$.
The probability of (I,C,I,I) is: $\frac{2}{3} \cdot \frac{1}{4} \cdot \frac{1}{2} \cdot \frac{4}{5} = \frac{8}{120}$.

The probability of (I,I,C,I) is: $\frac{2}{3} \cdot \frac{3}{4} \cdot \frac{1}{2} \cdot \frac{4}{5} = \frac{24}{120}$.

The probability of (I,I,I,C) is: $\frac{2}{3} \cdot \frac{3}{4} \cdot \frac{1}{2} \cdot \frac{1}{5} = \frac{6}{120}$.

So the probability of the event is $\frac{12+8+24+6}{120} = \frac{50}{120} = \frac{5}{12}$.

e) Zelda gets exactly half of the questions right.

Solution: There are six outcomes in which Zelda gets half of the questions right: (C,C,I,I), (C,I,C,I), (C,I,I,C), (I,C,C,I), (I,C,I,C), (I,I,C,C). The probability of each is:

- (C,C,I,I): $\frac{1}{3} \cdot \frac{1}{4} \cdot \frac{1}{2} \cdot \frac{4}{5} = \frac{4}{120}$
- (C,I,C,I): $\frac{1}{3} \cdot \frac{3}{4} \cdot \frac{1}{2} \cdot \frac{4}{5} = \frac{12}{120}$
- (C,I,I,C): $\frac{1}{3} \cdot \frac{3}{4} \cdot \frac{1}{2} \cdot \frac{1}{5} = \frac{3}{120}$
- (I,C,C,I): $\frac{2}{3} \cdot \frac{1}{4} \cdot \frac{1}{2} \cdot \frac{4}{5} = \frac{8}{120}$
- (I,C,I,C): $\frac{2}{3} \cdot \frac{1}{4} \cdot \frac{1}{2} \cdot \frac{1}{5} = \frac{2}{120}$
- (I,I,C,C): $\frac{2}{3} \cdot \frac{3}{4} \cdot \frac{1}{2} \cdot \frac{1}{5} = \frac{6}{120}$

Since these are disjoint, the total probability is:

$$\frac{4+12+3+8+2+6}{120} = \frac{35}{120} = \frac{7}{24}$$

f) the first question Zelda gets right is Question 3.

Solution: This event contains two outcomes: (I, I, C, I) and (I, I, C, C). The probability of the first is $\frac{2}{3} \cdot \frac{3}{4} \cdot \frac{1}{2} \cdot \frac{4}{5} = \frac{24}{120}$. The probability of the second is $\frac{2}{3} \cdot \frac{3}{4} \cdot \frac{1}{2} \cdot \frac{1}{5} = \frac{6}{120}$.

In total, the probability of the event is $\frac{30}{120} = \frac{1}{4}$.

Problem 4.

Out of one hundred students who applied to each of UC San Diego, UC Irvine, and UC Riverside:

- 35 were accepted to UCSD,
- 38 were accepted to Irvine, and
- 65 were accepted to Riverside.

Furthermore, 22 of these students were admitted to both UCSD and Irvine, 30 were admitted to both UCSD and Riverside, and 33 were admitted to both Irvine and Riverside.

Consider selecting a random student from the 100 who applied to each school, and let A be the even that the student was accepted to UCSD, let B be the event that the student was accepted to Irvine, and let C be the even that the student was accepted to Riverside.

For each event below, either compute its probability if there is enough information to do so, or state that there is not enough information.

a) $P(A \text{ or } B)$

Solution: This is computed by the formula:

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

We have all of these quantities:

$$\begin{aligned} &= \frac{35}{100} + \frac{38}{100} - \frac{22}{100} \\ &= \frac{51}{100} \end{aligned}$$

b) $P(A \text{ and } B)$

Solution: This is given: it is $22/100 = 0.22$.

c) $P(A \text{ or } C)$

Solution: This is computed by the formula:

$$P(A \text{ or } C) = P(A) + P(C) - P(A \text{ and } C)$$

We have all of these quantities:

$$\begin{aligned} &= \frac{35}{100} + \frac{65}{100} - \frac{30}{100} \\ &= \frac{70}{100} \end{aligned}$$

d) $P(A \text{ and } B \text{ and } C)$

Solution: We can't compute this with the information provided. This can be confirmed by drawing a Venn diagram.

e) $P(A \text{ or } B \text{ or } C)$

Solution: We can't compute this with the information provided. To do so, we would need $P(A \text{ and } B \text{ and } C)$. This can be confirmed by drawing a Venn diagram.

f) $P(A \text{ and } (B \text{ or } C))$

Solution: We can't compute this with the information provided. This can be confirmed by drawing a Venn diagram.