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

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

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.

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.

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

c) Zelda gets all of the questions right.

d) Zelda gets none of the questions right.

e) Zelda gets at least one question right.

f) Zelda gets exactly one of the questions right.

g) Zelda gets exactly half of the questions right.

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

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.

b) Zelda gets none of the questions right.

c) Zelda gets at least one question right.

d) Zelda gets exactly one of the questions right.

e) Zelda gets exactly half of the questions right.

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

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

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

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

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

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

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