22417.202210 Homework 2 - Linear Models for Classification

Pankaj Kumar Jatav

TOTAL POINTS

33 / 41

QUESTION 1

1 Math 1.1 **2 / 2**

✓ **- 0 pts** *Correct*

- 1 pts Partially correct

- 2 pts Incorrect

QUESTION 2

2 Math 1.2 **6 / 6**

✓ **- 0 pts** *Correct*

- 1 pts Almost correct

- 6 pts Incorrect

QUESTION 3

3 Math 1.3 **8 / 8**

✓ **- 0 pts** *Correct*

- 1 pts Almost correct

- 4 pts Partially correct

- 8 pts Incorrect

QUESTION 4

4 Perceptron Q.4 **1 / 1**

✓ **- 0 pts** *Correct*

QUESTION 5

5 Perceptron Q.5 **3 / 3**

✓ **- 0 pts** *Correct*

- 1 pts Miss one answer or give more answers

- 3 pts Incorrect

QUESTION 6

6 Perceptron Q.6 **2 / 2**

✓ **- 0 pts** *Correct*

- 2 pts Incorrect

QUESTION 7

7 Perceptron Q.7 **0 / 2**

- 0 pts Correct

- 1 pts Give two answers (one involves correct one)

✓ **- 2 pts** *Incorrect*

QUESTION 8

8 Perceptron Q.8 **2 / 2**

✓ **- 0 pts** *Correct*

QUESTION 9

9 Perceptron Q.9 **0 / 2**

- 0 pts Correct

- 1 pts Give two answers (one involves correct one)

✓ **- 2 pts** *Incorrect (give three answers or totally missed)*

QUESTION 10

10 Perceptron Q.10 **2 / 2**

✓ **- 0 pts** *Correct*

- 2 pts Incorrect

QUESTION 11

11 Perceptron Q.11 **2 / 2**

✓ **- 0 pts** *Correct*

- 2 pts Incorrect

QUESTION 12

9 pts

12.1 Perceptron Q.12 **0 / 2**

- 0 pts Correct

✓ **- 2 pts** *Incorrect*

12.2 Perceptron Q.12.b **0 / 2**

- 0 pts Correct

✓ **- 2 pts** *Incorrect*

12.3 Perceptron Q.12.c **1 / 1**

✓ **- 0 pts** *Correct*

- 1 pts Incorrect

12.4 Perceptron Q.12.d **3 / 3**

✓ **- 0 pts** *Correct*

- 3 pts Incorrect

12.5 Perceptron Q.12.e **1 / 1**

✓ **- 0 pts** *Correct*

- 1 pts Incorrect

QUESTION 13

13 Collaboration Questions **0 / 0**

✓ **- 0 pts** *Correct*

HOMEWORK 2
LINEAR MODELS FOR CLASSIFICATION[1]

CS 688 MACHINE LEARNING (SPRING 2022)
https://nlp.cs.gmu.edu/course/cs688-spring22/
Note: this assignment was partly created by Matt Gormley

OUT: Feb 3, 2022
DUE: Feb 9, 2022

Your name: Pankaj Kumar Jatav

Your GID: G01338769

[1]Compiled on Saturday 19th February, 2022 at 22:42

1 Written Questions [50 pts]

1.1 A bit of math

1. (2 points) Consider two non-negative numbers a and b, and show that, if $a \leq b$, then $a \leq (ab)^{1/2}$.

Work

Given that

$$a \leq b$$

Multiply both side with a, as a is non-negative number the inequality sign will remain same.

$$a * a \leq b * a$$

Taking the square root on both side

$$a \leq (ba)^{1/2}$$

2. (6 points) Use the above result to show that, if the decision regions of a two-class classification problem are chosen to minimize the probability of misclassification, this probability will satisfy:

$$p(\text{mistake}) \leq \int \{p(\mathbf{x}, \mathcal{C}_1)p(\mathbf{x}, \mathcal{C}_2)\}^{1/2} d\mathbf{x}.$$

Work

$$p(mistake) =\leq \int_{R1} p(\mathbf{x}, \mathcal{C}_2) d\mathbf{x} + \int_{R2} p(\mathbf{x}, \mathcal{C}_1) d\mathbf{x} ----(i)$$

In the error made first integral R1 We always have $p(C_1|x) \geq p(C_2|x)$. Then,

$$p(\mathcal{C}_2|x) \leq \{p(\mathcal{C}_1|x)p(\mathcal{C}_2|x)\}^{1/2}$$

$$\int_{R1} p(\mathbf{x}, \mathcal{C}_2) d\mathbf{x} = \int_{R1} p(\mathcal{C}_2|x)p(x) d\mathbf{x}$$

$$\leq \int_{R1} \{p(\mathcal{C}_1|x)p(\mathcal{C}_2|x)\}^{1/2} p(x) d\mathbf{x}$$

$$\leq \int_{R1} \{p(\mathcal{C}_1|x)p(\mathcal{C}_2|x)\}^{1/2} d\mathbf{x} ----- (ii)$$

Same apply for other part and we will have

$$\int_{R2} p(\mathbf{x}, \mathcal{C}_1) d\mathbf{x} \leq \int_{R2} \{p(\mathcal{C}_1|x)p(\mathcal{C}_2|x)\}^{1/2} d\mathbf{x} ----- (iii)$$

Using (i), (ii), (iii) we can say that

$$p(mistake) \leq \int_{R1} \{p(\mathcal{C}_1|x)p(\mathcal{C}_2|x)\}^{1/2} d\mathbf{x} + \int_{R2} \{p(\mathcal{C}_1|x)p(\mathcal{C}_2|x)\}^{1/2} d\mathbf{x}$$

$$p(\text{mistake}) \leq \int \{p(\mathbf{x}, \mathcal{C}_1)p(\mathbf{x}, \mathcal{C}_2)\}^{1/2} d\mathbf{x}$$

holds

1.2 Linear Discriminant Analysis

3. (8 points) Show that maximization of the class separation criterion given by (Bishop 4.22)

$$m_2 - m_1 = \mathbf{w}^T(\mathbf{m}_2 - \mathbf{m}_1)$$

with respect to $\mathbf{w}$, using a Lagrange multiplier to enforce the constraint $\mathbf{w}^T\mathbf{w} = 1$ leads to the result that $\mathbf{w} \propto (\mathbf{m}_2 - \mathbf{m}_1)$.

Work

$$m_2 - m_1 = \mathbf{w}^T(\mathbf{m}_2 - \mathbf{m}_1)$$

With a Lagrange multiplier we have,

$$f(w, \lambda) = m_2 - m_1 = \mathbf{w}^T(\mathbf{m}_2 - \mathbf{m}_1) + \lambda(1 - \mathbf{w}^T\mathbf{w})$$

$$f(w, \lambda) = \mathbf{w}^T(\mathbf{m}_2 - \mathbf{m}_1) + \lambda - \lambda\mathbf{w}^T\mathbf{w} - - - - - - -(i)$$

To find the maximization with respect to $\mathbf{w}$, Taking the gradient of above eq^n with respect to $\mathbf{w}$ and setting the gradient to zero

$$\frac{\partial f(\mathbf{w}, \lambda)}{\partial \mathbf{w}} = (\mathbf{m}_2 - \mathbf{m}_1) + 0 - 2\lambda\mathbf{w}$$

$$0 = (\mathbf{m}_2 - \mathbf{m}_1) - 2\lambda\mathbf{w}$$

$$2\lambda\mathbf{w} = (\mathbf{m}_2 - \mathbf{m}_1)$$

$$\mathbf{w} = \frac{(\mathbf{m}_2 - \mathbf{m}_1)}{2\lambda} - - - - - - -(ii)$$

Now, in the eq^n (ii) we do not know the max value of λ
, So To find the maximization with respect to λ, Taking the gradient of above $eq^n(i)$ with respect to λ and setting the gradient to zero

$$\frac{\partial f(\mathbf{w}, \lambda)}{\partial \lambda} = 0 + 1 - \mathbf{w}^T\mathbf{w}$$

$$0 = 1 - \mathbf{w}^T\mathbf{w}$$

Now replacing the $\mathbf{w}$ using $eq^n(ii)$

$$1 = \frac{(\mathbf{m}_2 - \mathbf{m}_1)}{2\lambda}^T \frac{(\mathbf{m}_2 - \mathbf{m}_1)}{2\lambda}$$

$$4\lambda^2 = (\mathbf{m}_2 - \mathbf{m}_1)^T(\mathbf{m}_2 - \mathbf{m}_1)$$

$$\lambda = \frac{||\mathbf{m}_2 - \mathbf{m}_1||}{2} - - - - - - -(iii)$$

Work

Replacing the λ value from $eq^n(iii)$ in $eq^n(ii)$

$$\mathbf{w} = \frac{(\mathbf{m}_2 - \mathbf{m}_1)}{||\mathbf{m}_2 - \mathbf{m}_1||}$$

$||\mathbf{m}_2 - \mathbf{m}_1||$ will be always non negative value so, it can be written as.

$$\mathbf{w} \propto (\mathbf{m}_2 - \mathbf{m}_1)$$

2 Perceptron

4. (1 point) Consider running the online perceptron algorithm on some sequence of examples S (an example is a data point and its label). Let S' be the same set of examples as S, but presented in a different order.

   True or False: the online perceptron algorithm is guaranteed to make the same number of mistakes on S as it does on S'.

   Select one:

   ● True

   ○ False

5. (3 points) Suppose we have a perceptron whose inputs are 2-dimensional vectors and each feature vector component is either 0 or 1, i.e., $x_i \in \{0, 1\}$. The prediction function $y = \text{sign}(w_1 x_1 + w_2 x_2 + b)$, and

$$\text{sign}(z) = \begin{cases} 1, & \text{if } z \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

   Which of the following functions can be implemented with the above perceptron? That is, for which of the following functions does there exist a set of parameters w, b that correctly define the function. **Select all that apply:**

   ■ AND function, i.e., the function that evaluates to 1 if and only if all inputs are 1, and 0 otherwise.

   ■ OR function, i.e., the function that evaluates to 1 if and only if at least one of the inputs are 1, and 0 otherwise.

   □ XOR function, i.e., the function that evaluates to 1 if and only if the inputs are not all the same. For example

$$\text{XOR}(1, 0) = 1, \text{ but } \text{XOR}(1, 1) = 0.$$

   □ None of the above.

6. (2 points) Suppose we have a dataset $\{(\mathbf{x}^{(1)}, y^{(1)}), \ldots, (\mathbf{x}^{(N)}, y^{(N)})\}$, where $\mathbf{x}^{(i)} \in \mathbb{R}^M$, $y^{(i)} \in \{+1, -1\}$. We would like to apply the perceptron algorithm on this dataset. Assume there is no intercept (bias) term. How many parameter values is the perceptron algorithm learning?

   Select one:

   ○ N

   ○ $N \times M$

   ● M

7. (2 points) Which of the following are true about the perceptron algorithm?

   Select all that apply:

   - ■ The number of mistakes the perceptron algorithm makes is proportional to the number of points in the dataset.
   - □ The perceptron algorithm converges on any dataset.
   - ■ The perceptron algorithm can be used in the context of online learning.
   - ■ For linearly separable data, the perceptron algorithm always finds the separating hyperplane with the largest margin.
   - □ None of the above.

8. (2 points) The following table shows a data set and the number of times each point is misclassified during a run of the perceptron algorithm. What is the separating plane w found by the algorithm, i.e. $w = [b, w_1, w_2, w_3]$? Assume that the initial weights are all zero.

x_1	x_2	x_3	y	Times Misclassified
2	1	5	1	10
5	3	3	1	5
4	1	2	1	8
8	4	8	-1	2
3	2	6	-1	3

   Select one:

   - ○ $[3, 22, 11, 24]$
   - ○ $[0, -5, 20, -10]$
   - ○ $[16, 56, 18, 47]$
   - ● $[18, 52, 19, 47]$

9. (2 points) Please select the correct statement(s) about the mistake bound of the perceptron algorithm.

   Select all that apply:

   - □ If the minimum distance from any data point to the separating hyperplane is increased, without any other change to the data points, the mistake bound will also increase.
   - □ If the whole dataset is shifted away from origin, then the mistake bound will also increase.
   - ■ If the size of the data set (i.e., the maximum pair-wise distance between data points) is increased, then the mistake bound will also increase.
   - ■ The mistake bound is linearly inverse-proportional to the minimum distance of any data point to the separating hyperplane of the data.
   - □ None of the above.

10. (2 points) Suppose we have data whose elements are of the form $[x_1, x_2]$, where $x_1 - x_2 = 0$. We do not know the label for each element. Suppose the perceptron algorithm starts with $\theta = [3, 5]$, which of the following values will θ never take on in the process of running the perceptron algorithm on the data?

    Select one:

    ○ $[-1, 1]$

    ○ $[4, 6]$

    ● $[-3, 0]$

    ○ $[-6, -4]$

11. (2 points) Consider the linear decision boundary below and the training dataset shown. Which of the following are valid weights θ and its corresponding training error? (Note: Assume the decision boundary is fixed and does not change while evaluating training error.)

    Select all that apply:

    ■ $\theta = [2, 1]$, error = 5/13

    □ $\theta = [-2, 1]$, error = 5/13

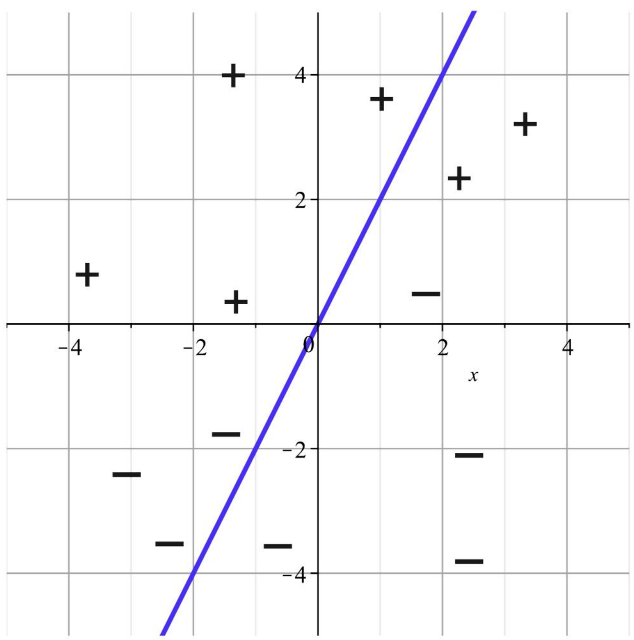
    □ $\theta = [2, -1]$, error = 8/13

    □ $\theta = [2, -1]$, error = 5/13

    □ $\theta = [-2, 1]$, error = 8/13

    ■ $\theta = [-2, -1]$, error = 8/13

    □ None of the above.



12. (9 points) The following problem will walk you through an application of the Perceptron Mistake Bound. The following table shows a linearly separable dataset. Your task will be to determine the mistake bound for the dataset below. Then, you will run the training of the perceptron.

x_1	x_2	y
-1.939	2.704	1
-0.928	-3.054	-1
-2.181	-3.353	-1
-0.142	1.440	1
2.605	-0.651	1

(a) (2 points) Compute the radius R of the "circle" which bounds the data points. Round your answer to 3 decimal places after the decimal point.

Radius:

2.9606

(b) (2 points) Assume that the linear separator with the largest margin is given by

$$\boldsymbol{\theta}^{*T} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = 0, \text{, where } \boldsymbol{\theta}^* = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}$$

Now, compute the margin of the dataset. Round your answer to 3 decimal places after the decimal point.

Margin:

1.841384

(c) (1 point) Based on the above values, what is the theoretical perceptron mistake bound for this dataset, given this linear separator? Round your answer to 3 decimal places after the decimal point.

Mistake Bound:

2.5850

(d) (3 points) Finally, implement and train the perceptron using the dataset. Train using the datapoints **in the given order, top to bottom** repeatedly until convergence. Give the final weights in a comma-separated list (w_1, w_2, b).

Weights:

(7.066, 4.806, 0)

(e) (1 point) How many mistakes did the perceptron make?

Mistakes:

2

Collaboration Questions Please answer the following:

1. Did you receive any help whatsoever from anyone in solving this assignment?
   No.
   - If you answered 'yes', give full details: ____________________
   - (e.g. "Jane Doe explained to me what is asked in Question 3.4")
2. Did you give any help whatsoever to anyone in solving this assignment?
   No.
   - If you answered 'yes', give full details: ____________________
   - (e.g. "I pointed Joe Smith to section 2.3 since he didn't know how to proceed with Question 2")
3. Did you find or come across code that implements any part of this assignment ?
   No. (See below policy on "found code")
   - If you answered 'yes', give full details: ____________________
   - (book & page, URL & location within the page, etc.).