First Name: ____________________ Last Name: ____________________

Student Number: ____________________

ECE 421S — Introduction to Machine Learning

Midterm Examination
Monday, Febrary 24^{th}, 2020
6:15 p.m. – 8:00 p.m.

Instructors: Ashish Khisti and Ben Liang

Circle your tutorial section:

- TUT0101 Thu 3-5 (SF2202)
- TUT0102 Thu 3-5 (GB304)
- TUT0103 Tue 10-12 (SF2202)
- TUT0104 Fri 9-11 (BA1230)

Instructions

- Please read the following instructions carefully.
- You have one hour forty-five minutes (1:45) to complete the exam.
- Please make sure that you have a complete exam booklet.
- Please answer *all* questions. Read each question carefully.
- The value of each question is indicated. Allocate your time wisely!
- No additional pages will be collected beyond this answer book. You may use the reverse side of each page if needed to show additional work.
- This examination is closed-book; One 8.5 × 11 aid-sheet is permitted. A non-programmable calculator is also allowed.
- Good luck!

1. (20 MARKS) Consider a binary linear classification problem where the data points are two dimensional, i.e., $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$ and the labels $y \in \{-1, +1\}$. Throughout this problem consider a data-set with the following four points:

$$\mathcal{D} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), (\mathbf{x}_3, y_3), (\mathbf{x}_4, y_4)\}$$

where the input data-vectors are given by

$$\mathbf{x}_1 = (0,0)^T, \quad \mathbf{x}_2 = (0,1)^T, \quad \mathbf{x}_3 = (1,1)^T, \quad \mathbf{x}_4 = (1,0)^T.$$

and the associated labels are given by

$$y_1 = +1, \quad y_2 = -1, \quad y_3 = +1, \quad y_4 = -1.$$

Our aim is to find a linear classification rule $w_0 + w_1 x_1 + w_2 x_2$, with weight vector $\mathbf{w} = (w_0, w_1, w_2)^T$, that classifies this dataset.

[**Important**: Recall that, in our learning algorithms and their analysis, we transform the data vectors to include the constant term, i.e., $\mathbf{x}_1 = (0,0)^T$ must be transformed to $\tilde{\mathbf{x}}_1 = (1,0,0)^T$ etc.]

5 marks (**a**) Is there a weight vector $\mathbf{w}$ that satisfies the following property?

$$y_n(\mathbf{w}^T \mathbf{x}_n) > 0 \quad \text{for all } n \in \{1,2,3,4\}.$$

If your answer is yes, find such $\mathbf{w}$. If your answer is no, explain why not.

total/5

10 marks

(**b**) Suppose we implement the perceptron learning algorithm as discussed in class with the initial weight vector $\mathbf{w} = (3.5, 0, 1)^T$ and the standard update rule for mis-classified points. Assume that each point that falls on the boundary is treated as a mis-classified point. The algorithm visits the points in the following order:

$$\mathbf{x}_1 \to \mathbf{x}_2 \to \mathbf{x}_3 \to \mathbf{x}_4 \to \mathbf{x}_1 \to \mathbf{x}_2 \cdots$$

until the training error $E_{\text{in}}(\mathbf{w}) \leq \frac{1}{4}$, at which time the algorithm terminates. Here the training error is defined as usual:

$$E_{\text{in}}(\mathbf{w}) = \frac{1}{4}\sum_{n=1}^{4} \mathbf{1}(\hat{y}_n \neq y_n),$$

where $\mathbf{1}(\cdot)$ is the indicator function and $\hat{y}_n$ is the output of the classifier.

Show the output of the perceptron learning algorithm in each step and sketch the final decision boundary when the algorithm terminates.

total/-

[continue part (b) here]

total/10

5 marks

(c) Suppose we now consider binary logistic regression to classify the points in $\mathcal{D}$, with the following sigmoid function for likelihood:

$$\hat{P}(y = +1|\mathbf{x}) = \theta(\mathbf{w}^T\mathbf{x}) = \frac{e^{\mathbf{w}^T\mathbf{x}}}{1 + e^{\mathbf{w}^T\mathbf{x}}}.$$

Assume that we use the log-loss function (i.e., log-likelihood) to measure training error as discussed in class. Among the following two possible solutions for $\mathbf{w}$, which one would we prefer?

$$\mathbf{w}_1 = (0.5, -1, -1)^T$$
$$\mathbf{w}_2 = (0.5, -1, 0)^T$$

total/5

2. (20 MARKS) Consider linear regression over dataset $\mathcal{D} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_N, y_N)\}$. Suppose the labels $y_n \in \{-1, +1\}$, for $n = 1, 2, \ldots, N$. For any given model parameter $\mathbf{w}$, instead of the usual squared error, we use the following loss function for each example $\mathbf{x}_n$:

$$e_n(\mathbf{w}) = \left(\max(0, 1 - y_n \mathbf{w}^T \mathbf{x}_n)\right)^2.$$

The training error $E_{\text{in}}(\mathbf{w}) = \frac{1}{N} \sum_{n=1}^{N} e_n(\mathbf{w})$.

10 marks (**a**) Derive the gradient $\nabla E_{\text{in}}(\mathbf{w})$.

total/10

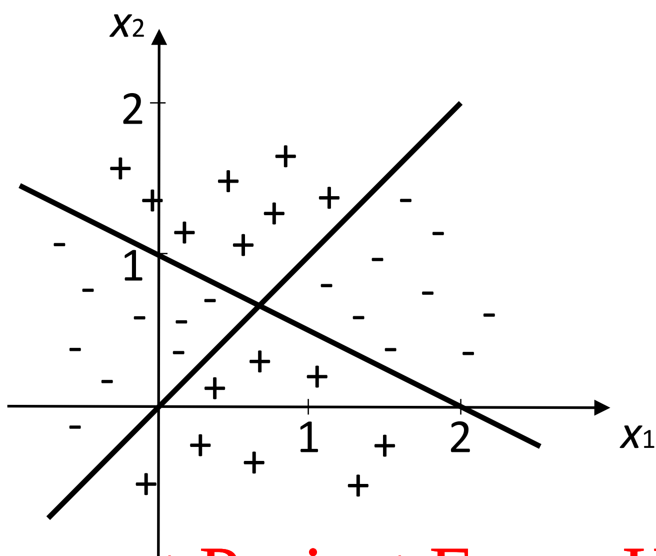
5 marks (b) Write the pseudo code for the Stochastic Gradient Descent method to minimize $E_{\text{in}}(\mathbf{w})$, with minibatch size 1.

total/5

5 marks (c) What does it mean to have $e_n(\mathbf{w}) = 0$ for some given $\mathbf{w}$ and $\mathbf{x}_n$? You should give a **geometric** interpretation. (Hint: consider two cases depending on the value of y_n.)

total/5

3. (20 MARKS) Consider a binary linear classification problem where the data points are two dimensional, i.e., $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$ and the labels $y \in \{-1, +1\}$. We wish to build a multi-layer perceptron to classify the dataset as shown below, where the "+" and "-" signs indicate examples with labels +1 and −1, respectively.



6 marks

(**a**) Design two perceptrons to implement the two lines shown in the figure above. The lines are $x_2 = x_1$ and $x_2 = -\frac{1}{2}x_1 + 1$. To ensure uniformity for easy marking by the teaching staff, each of your perceptrons **must** classify the region **above** the line to +1.

total/6

6 marks

(**b**) Design a (**single-layer**) perceptron to implement the NAND function. Recall that $\mathrm{NAND}(a, b) = \mathrm{NOT}(\mathrm{AND}(a, b)) = \mathrm{OR}(\mathrm{NOT}(a), \mathrm{NOT}(b))$, where a and b are binary variables. You should use $\{-1, +1\}$ to label a binary variable as shown in class.

total/6

8 marks (**c**) Use **only** the perceptrons in parts (a) and (b) to build a multi-layer perceptron to classify the dataset in the figure. You may use as many copies of these perceptrons as you need. **Draw** your design and clearly label all edges and weights.

total/8