FIRST NAME: ______________________ LAST NAME: ______________________

STUDENT NUMBER: ______________________

ECE 521S — Inference Algorithms and Machine Learning

Test 1

February 7^{th}, 2018
4:15 p.m. – 4:55 p.m.

Instructors: Stark Draper and Ashish Khisti

Grading Scheme

Instructions

- Please read the following instructions carefully.
- You have forty (0:40) 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 the 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 (double sided) is permitted. A non-programmable calculator is also allowed.
- Good luck!

1. (10 MARKS) Consider a binary linear classification problem, where the data points are two dimensional, i.e., $\mathbf{x} \in \mathbb{R}^2$, and the labels $y \in \{-1, 1\}$. The training set consists of the following points as shown in the figure below.

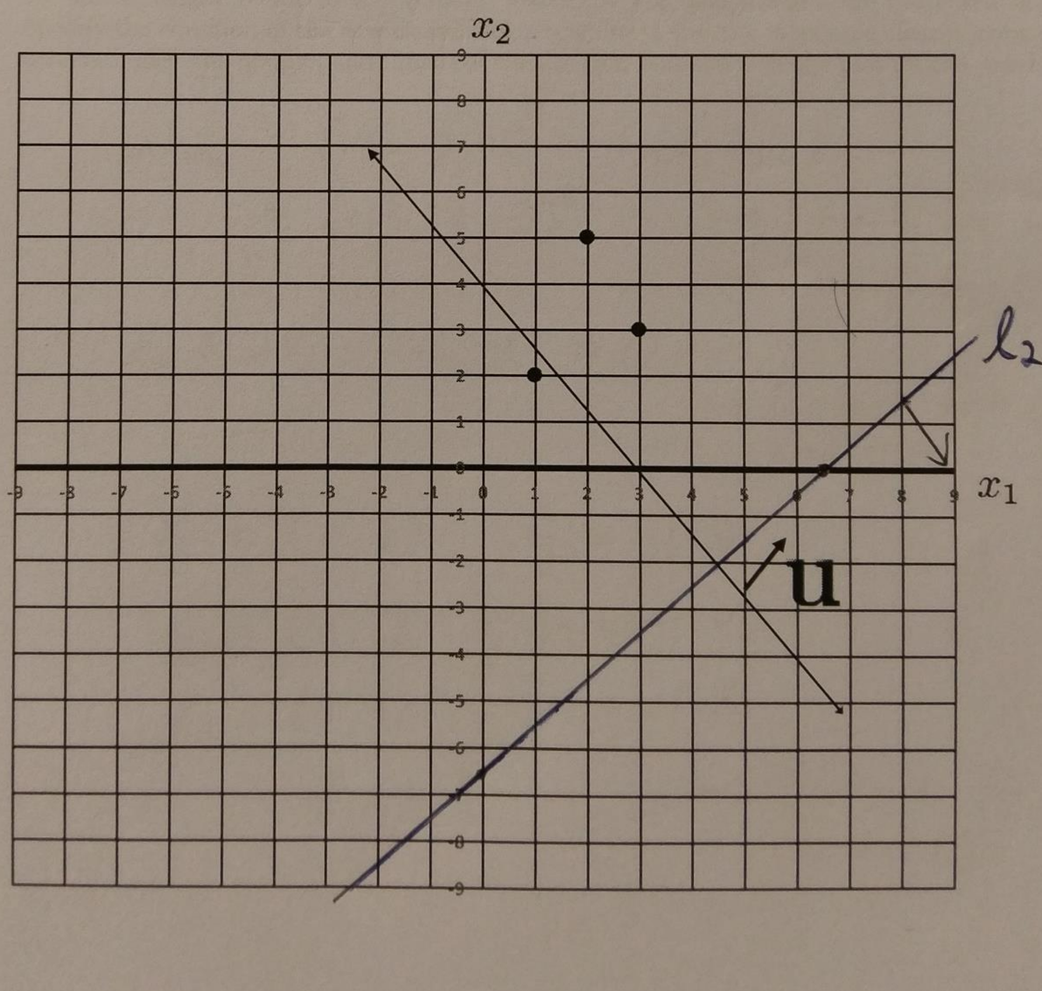
$$\mathbf{x}_1 = (1,2)^T, \qquad y_1 = -1$$
$$\mathbf{x}_2 = (2,5)^T, \qquad y_2 = -1$$
$$\mathbf{x}_3 = (3,3)^T, \qquad y_3 = +1$$

Suppose we wish to implement a linear classifier using the Perceptron Learning algorithm:

$$y = \text{sign}(w_0 + w_1 x_1 + w_2 x_2)$$

Suppose that the perceptron algorithm is initialized to the straight line in the figure below. This line has an intercept of 3 along the horizontal axis and 4 along the vertical axis.

All points above this line i.e., in the direction pointed by vector u are labeled as +1 whereas all points below the line i.e., in the direction pointed by −u are labelled as −1.



~~2~~ marks 3

(a) Provide an equation of the straight line that determines the decision boundary in the above figure. Express your answer in the form: $ax_1 + bx_2 + c = 0$ and specify what the constants a and b are assuming that we fix $c = -12$.

$$4x_1 + 3x_2 - 12 = 0$$

→ 1.5 marks for $a=4$

→ 1.5 marks for $b=3$

2 marks

(b) Provide the in-sample classification error E_{in} associated with the decision boundary in part (a).

$$E_{in} = \frac{1}{3}$$

→ either full or zero grade

~~6~~ marks 5

(c) Perform an update of the perceptron algorithm with respect to the mis-classified point by taking the initial weight vector to $\mathbf{w} = (c, a, b)$, where $c = -12$, and a and b are evaluated in part (a). Specify the equation of the new classification boundary, find the in-sample classification error E_{in} achieved after the update, and illustrate the decision boundary on the plot on the previous page.

$$W_{new} = (-12, 4, 3)^T + (-1)(1, 2, 5)^T$$
$$= (-13, 2, -2)^T$$

→ 1 mark for $(-12,4,3)^T$

→ 1 mark for $y_2\underline{x}_2 = (-1)(1,2,5)^T$

$$l_2 : 2x_1 - 2x_2 - 13 = 0$$

→ 1 mark for correct equation and plot

$\underline{x}_1 = (1, 2)$ $\hat{y}_1 = -1$ ✓

$\underline{x}_2 = (2, 5)$ $\hat{y}_2 = -1$ ✓

$x_3 = (3, 3)$ $\hat{y}_3 = -1$ X

$$E_{in} = \frac{1}{3}$$

→ 2 marks for finding the correct E_{in}

total/10

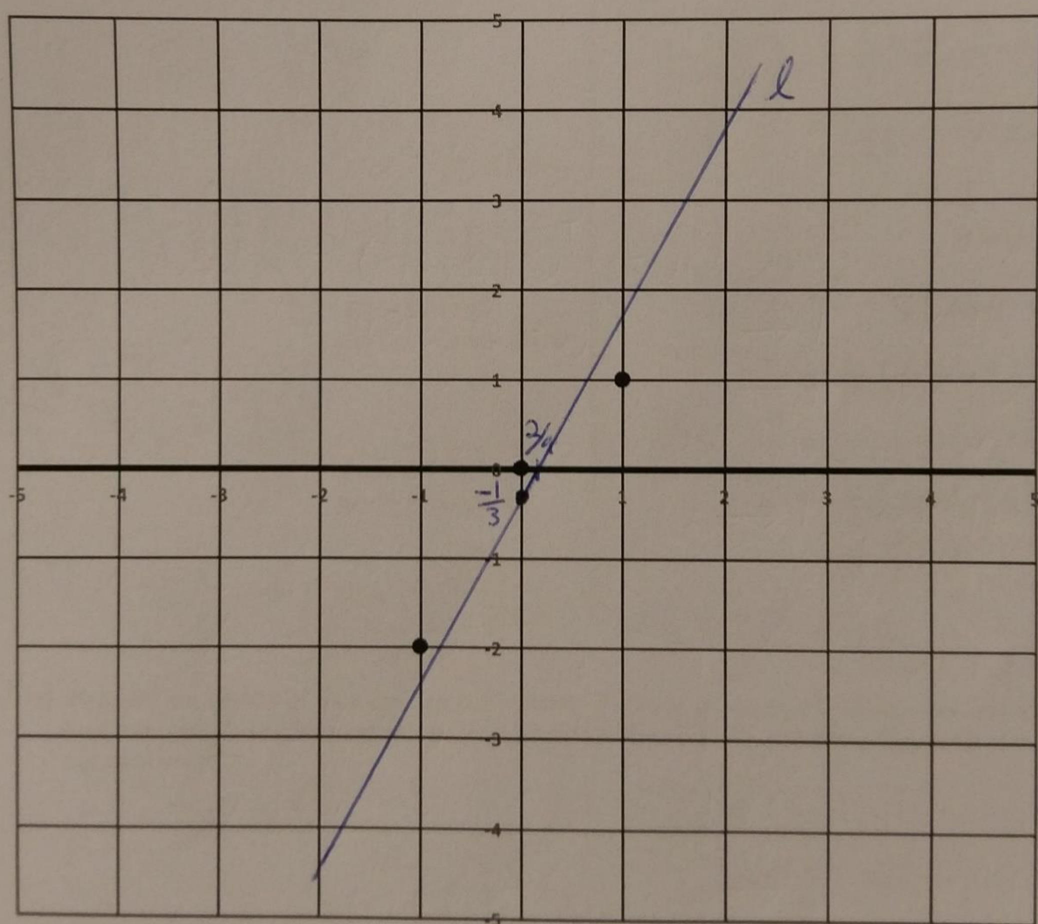
2. (10 MARKS) Suppose that we are given three points in the (x, y) plane, with the following coordinates:

$$x_1 = -1, \quad y_1 = -2,$$
$$x_2 = 0, \quad y_2 = 0,$$
$$x_3 = 1, \quad y_3 = 1.$$

We wish to find a straight line approximation: $\hat{y} = w_0 + w_1 x$ that minimizes:

$$E_{\text{in}}(w_0, w_1) = \frac{1}{3}\sum_{i=1}^{3}(y_i - \hat{y}_i)^2,$$

where $\hat{y}_i = w_0 + w_1 x_i$ is the estimate of y_i based on x_i on the straight line.



total/10

Please answer the following questions on the next page:

3 marks

(a) Rewrite the above problem specifications based on x_i to get the problem into the standard form of a least squares problem: $\mathbf{w}^\star = \arg\min_{\mathbf{w}\in\mathbb{R}^2} \frac{1}{3}||\mathbf{X}\mathbf{w} - \mathbf{y}||^2$. Specifically specify the data matrix $\mathbf{X}$ and the target vector $\mathbf{y}$, where $\mathbf{w} = (w_0, w_1)^T$.

$$\mathbf{X} = \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} \qquad \underline{Y} = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$$

→ 1 mark for correct values of $\underline{Y}$

→ 2 marks for correct values of $\mathbf{X}$

4 marks

(b) Find the least squares solution $\mathbf{w}^\star = (w_0^\star, w_1^\star)$ in part (a). Sketch the line $\hat{y} = w_0^\star + w_1^\star x$ on the plot on the previous page.

Method 1:

$$W_{LS}^* = (X^T X)^{-1} X^T Y$$

→ 1 mark for correct equation

$$= \left(\begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$$

→ 1 mark for correct calculation

$$= \begin{bmatrix} 1/3 & 0 \\ 0 & 1/2 \end{bmatrix} \begin{bmatrix} -1 \\ 3 \end{bmatrix} = \begin{bmatrix} -1/3 \\ 3/2 \end{bmatrix}$$

→ 1 mark for correct answer

$$l: \hat{y} = \frac{-1}{3} + \frac{3}{2}x$$

→ 1 mark for correct equation and sketch

Method 2:

$$E_{in}(w^*) = \frac{1}{3}||XW - Y||^2$$

$$= \frac{1}{3}\left[(W_0 - W_1 + 2)^2 + W_0^2 + (W_0 + W_1 - 1)^2\right]$$

→ 1 mark for correct equation

$$\frac{\partial E_{in}(w^*)}{\partial W_0} = 0 = \frac{1}{3}\left[2(W_0 - W_1 + 2) + 2W_0 + 2(W_0 + W_1 - 1)\right]$$

$$= \frac{1}{3}(6W_0 + 2) \Rightarrow W_0 = \frac{-1}{3}$$

$$\frac{\partial E_{in}(w^*)}{\partial W_1} = 0 = \frac{1}{3}\left[2(W_0 - W_1 + 2)(-1) + 2(W_0 + W_1 - 1)\right]$$

$$= \frac{1}{3}(4W_1 - 6) \Rightarrow W_1 = \frac{3}{2}$$

→ 1 mark for calculation

→ 1 mark for correct answer

$$l: \hat{y} = \frac{-1}{3} + \frac{3}{2}x$$

→ 1 mark for equation and sketch

3 marks

(c) Suppose we decide to fit a quadratic function: $\hat{y} = w_0 + w_1 x + w_2 x^2$. What data-matrix $\mathbf{X}$ should you now use in the least squares formulation? What will be the error $E_{\text{in}}(\mathbf{w}^\star)$ attained by the optimal solution?

$$\mathbf{X} = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix} \qquad E_{in} = 0$$

→ 1 mark for correct values of $\mathbf{X}$

→ 2 marks for showing $E_{in} = 0$

total/10