Homework 3 (Optional) **Name: _____________**
ECE/CS 498 DS Spring 2020 **NetID: _____________**
Issued: 03/02/2020

This homework will not be graded. These problems are meant to provide you with some practice before your exam on 3/11/2020.

Problem 1

Given the following data points in the cartesian coordinate: {(*-1*, *0*), (*0*, *1*), (*2*, *4*)}

a) In this sub-question, you'll derive the least square linear regression expression in terms of slope and intercept (e.g. $y = mx + b$) for the 3 given data points from first principles. Please follow the steps below.

1. Fill in the table below. An example is given for the point (-1,0).

x	y	$y' = mx + b$	$y - y'$
-1	0	$m(-1) + b = -m + b$	$0 - (-m + b) = m - b$
0	1	b	$1 - b$
2	4	$2m + b$	$4 - 2m - b$

2. Provide the expression for summation of squared error (SSE), i.e. $SSE = \sum(y - y')^2$

$$(m - b)^2 + (1 - b)^2 + (4 - 2m - b)^2$$

3. Simplify the expression of SSE and express it as follows:
   i. $SSE = f_1(m)$, where $f_1(m)$ is a quadratic function of the variable m. Note that terms with b are part of the co-efficient or constant.
   ii. $SSE = f_2(b)$, where $f_2(b)$ is a quadratic function of the variable b. Note that terms with m are part of the co-efficient or constant.

i) $f(m) = 5m^2 + (2b - 16)m + 3b^2 - 10b + 17$
ii) $f(b) = 3b^2 + (2m - 10)b + 5m^2 - 16m + 17$

Homework 3 (Optional) **Name: ____________**
ECE/CS 498 DS Spring 2020 **NetID: ____________**
Issued: 03/02/2020

4. Find the *argmin* values for each quadratic function you derived above:
   i. Find $\underset{m}{\operatorname{argmin}} f_1(m)$. The answer will be an expression for m in terms of b.
   ii. Find $\underset{b}{\operatorname{argmin}} f_2(b)$. The answer will be an expression for b in terms of m.

   i. $m = \frac{8-b}{5}$
   ii. $b = \frac{5-m}{3}$

5. The answer from part 4 is a set of two linear equations. Solve the linear equations simultaneously for m and b. Write down the final mathematical expression of the linear regression line.

$$m = \frac{19}{14}, b = \frac{17}{14}$$
$$y = \frac{19}{14}x + \frac{17}{14}$$

b) The general formula for m and b are given below:

$$m = \frac{N(\sum_{i=1}^{N} x_i y_i) - (\sum_{i=1}^{N} x_i) \cdot (\sum_{i=1}^{N} y_i)}{N(\sum_{i=1}^{N} x_i^2) - (\sum_{i=1}^{N} x_i)^2}$$

$$b = \frac{(\sum_{i=1}^{N} y_i) \cdot (\sum_{i=1}^{N} x_i^2) - (\sum_{i=1}^{N} x_i) \cdot (\sum_{i=1}^{N} x_i y_i)}{N(\sum_{i=1}^{N} x_i^2) - (\sum_{i=1}^{N} x_i)^2}$$

Where N denotes the number of data points available to fit the linear regression line. In our case, $N = 3$.
Calculate m and b by plugging the corresponding numerical values in the formulae above. Are the results the same as the value of m and b calculated in a)?

$$b = \frac{5 \cdot 5 - 1 \cdot 8}{3 \cdot 5 - 1} = \frac{17}{14}$$
$$m = \frac{3 \cdot 8 - 1 \cdot 5}{3 \cdot 5 - 1} = \frac{19}{14}$$

Yes.

Homework 3 (Optional) **Name: ____________**
ECE/CS 498 DS Spring 2020 **NetID: ____________**
Issued: 03/02/2020

Problem 2

Consider a linear regression model, where each data point is represented by input x, target variable y with the following relationship:

$$y = w \cdot x + \epsilon$$

where w is a single real-valued parameter to be learned, and ϵ, the *noise* term, is independently and identically drawn from a Gaussian distribution with mean 0 and variance 1, i.e. $\epsilon \sim N(\mu = 0, \sigma^2 = 1)$.
Provide the mathematical expression for conditional probability $p(y|w,x)$ in terms of y, w, x.

$$p(\epsilon) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(\epsilon-\mu)^2}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi}} exp(-\frac{\epsilon^2}{2})$$

Plug in $y^{(i)} = w \cdot x^{(i)} + \epsilon^{(i)} \Rightarrow \epsilon^{(i)} = y^{(i)} - w \cdot x^{(i)}$, we obtain,

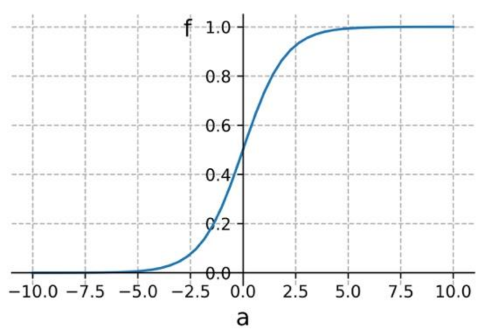
$$p\left(y^{(i)}|x^{(i)}, w\right) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}\left(y^{(i)} - w \cdot x^{(i)}\right)^2)$$

Problem 3

As we saw in lecture slides (L09 p36), in generalized linear model, $E[Y] = \mu = f(\beta^T X)$. Here *f(.)* is called an **activation function**. Recall that the lecture slides used link function which is the inverse of the activation function i.e., $g(.) = f^{-1}(.)$.

a) i) Write out the expression of the activation/sigmoid function $f(\cdot)$ that is used in logistic regression; ii) Plot $f(\cdot)$;

$$f(a) = \frac{1}{1 + e^{-a}}$$



b) How do we further map the output of the activation function to binary class labels $y \in \{-1,1\}$? [*Hint:* think in terms of *function*]

$sgn(f(\cdot) - 0.5) or\ 2u(f(\cdot) - 0.5) - 1$ (u is the step function)

Homework 3 (Optional)
ECE/CS 498 DS Spring 2020
Issued: 03/02/2020

Name: _____________
NetID: _____________

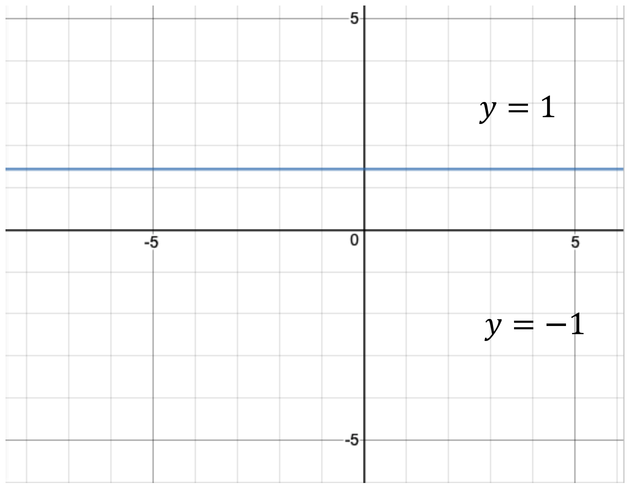
Problem 4

Draw the decision boundary and label values at both sides of the boundary for the following logistic regression classifier on the cartesian coordinate:

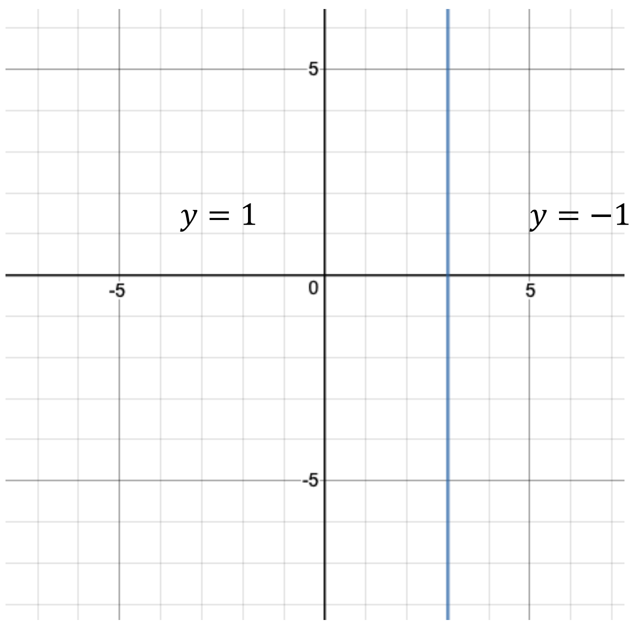
$$h_\theta(x) = sign(f(\theta_0 + \theta_1 x_1 + \theta_2 x_2) - 0.5)$$

a) $\theta_0 = -10, \theta_1 = 0, \theta_2 = 7$
b) $\theta_0 = 6, \theta_1 = -2, \theta_2 = 0$
c) $\theta_0 = 8, \theta_1 = 2, \theta_2 = 4$

a)

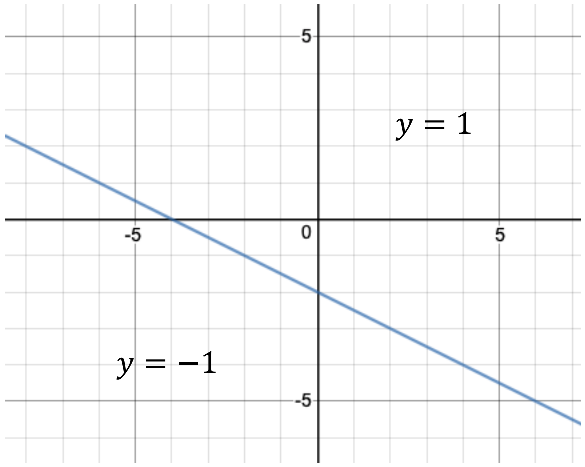


b)



c)

Homework 3 (Optional) **Name: ____________**
ECE/CS 498 DS Spring 2020 **NetID: ____________**
Issued: 03/02/2020



Problem 5

For *binary* target variable y, compare logistic regression and Naïve Bayes. Each input variable x has *k* binary features.

a) How many parameters are needed in Naïve Bayes model? *2k+1*
b) How many parameters are needed in logistic regression model? (Don't forget the bias term) *k+1*
c) Write out the conditional independence assumption in Naïve Bayes model.
   Given class label, features are conditionally independent
d) Write out one independence assumption in logistic regression model.
   Each data point is i.i.d. (independently and identically distributed)

Problem 6

As described in the lecture, one way of finding the first principal component, i.e., the eigenvector corresponding to the largest eigenvalue of a matrix is to consider an arbitrary vector and keep multiplying it with the matrix till the direction of vector doesn't change any more. This algorithm is known as *power iteration* (for details, refer to: https://en.wikipedia.org/wiki/Power_iteration). In this problem we will find the largest eigenvalue and the corresponding eigenvector **using Python.**

Consider the matrix, S = $\begin{bmatrix} 10 & 3 \\ 3 & 6 \end{bmatrix}$, and x = $\begin{bmatrix} 3 \\ 5 \end{bmatrix}$. Starting with the vector x, perform power iteration to find the largest eigenvector and eigenvalue. You can follow the steps given below:

i) $x_{new} = Sx_{old}$
ii) $x_{new} = \frac{x_{new}}{||x_{new}||}$
iii) Check if the x_{new} and x_{old} are the same i.e., the algorithm has converged. If they are, then terminate. If they are not, then go back to step i) and use x_{new} as x_{old}.

Homework 3 (Optional) **Name: ____________**
ECE/CS 498 DS Spring 2020 **NetID: ____________**
Issued: 03/02/2020

Please initialize appropriately. For this problem, you can declare that the algorithm has converged if the Euclidean distance between x_{new} and x_{old} is $<10^{-5}$. Answer the following questions:

1. How many iterations of the above algorithm does it require to converge to the first principal component?
   12
2. What is the first principal components i.e., eigenvector corresponding to the largest eigenvalue?
   [0.882, 0.472]
3. What is the largest eigenvalue?
   11.61
4. On the same plot, plot the vectors from all iterations. What do you observe? (Hint: Make sure to normalize each vector to be unit length for better visualization).