Com S 573	Spring Semester 2021
Machine Learning	
Homework 1	
Instructor: Hongyang Gao	Submitted by: Ibne Farabi Shihab

1 Problem 1:

1.a

Given that

$w = [w_0, w_1, w_2]^T$ and x= $[1, x_1, x_2]^T$

$$So, W^T = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix}$$

$$W^T.X = \begin{bmatrix} w_0 \\ w_1 \\ w_2 \end{bmatrix} [1, x_1, x_2]$$

After the multiplication we get: $w_0 * 1 + w_1 * X_1 + w_2 * X_2$

It is clearly visible that our intercept here is $b = w_0$

From there we can find x_2 which is our expected slope named α :

$x_2 = -(w_1/w_2) * x_1 – (w_0/w_2)$

1.b

Using $w = [1, 2, 3]^T$ we get,

$x_2 = -(2/3) * x_1 – (1/3)$

When we use $w = -[1, 2, 3]^T$ we get,

$x_2 = –(1/3) - (2/3) * x_1$

2 Question number 2:

2.a

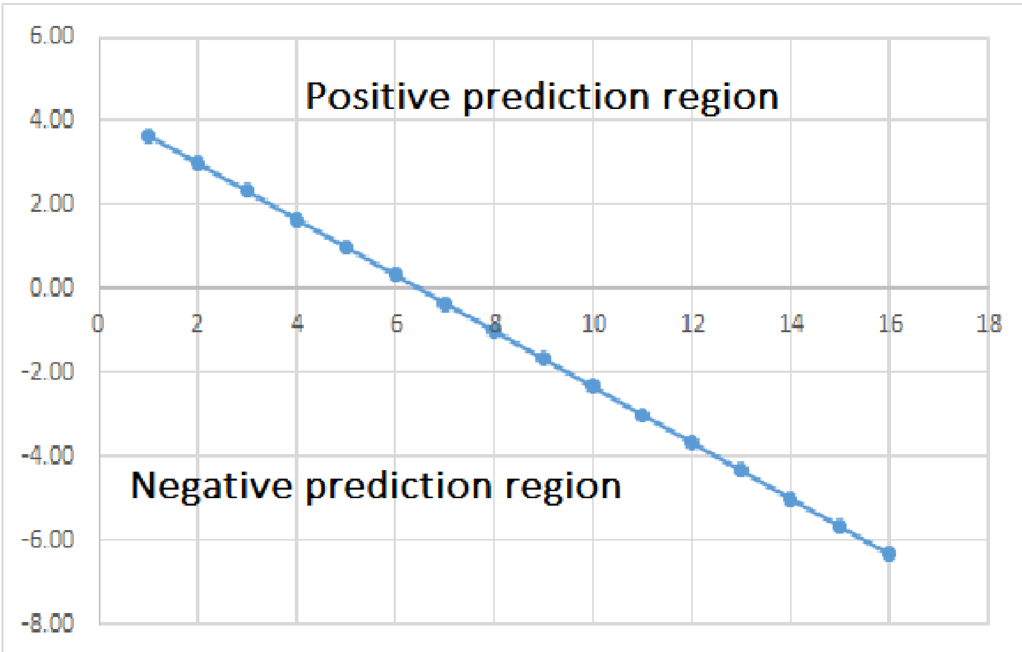

Figure 1: Question 1(b) graph

Given that:

$E(w) = \frac{1}{N} \sum ln(1 + e^{(} - y_n * w^T * x_n))$

$\frac{dE(w)}{dw} = -\frac{1}{N} \sum (1/(1 + exp^{(} - y_n * w^T * x_n)) \frac{d}{dw}(1 + exp^{(} - y_n * w^T * x_n))$

$= \frac{1}{N} \sum (y_n * x_n * exp^{(} - y_n * w^T * x_n))/(1/(1 + exp^{(} - y_n * w^T * x_n)))$

2.b

Here the input to the algorithm is linear combination of features. To be more specific our input's weight do not have any relation . For instance $= [x_0 * w_0 + x_1 * w_1, x_2 * *w_2]$ is a linear combination. Add to that, we decide on logistic regression based on a threshold value which makes the decision boundary linear.

If non-linear input features were used the decision boundary would be non-linear, For instance, $[x_0 * w_0 + x_1^2 * w_1, x_2^3 * *w_2]$

2.c

Changing threshold from 0.5 to 0.9 will move the decision boundary by 0.4, which does not make it non-linear.

2.d

As long as the input is a linear combination of features and the weights does not mix up with each other, logistic regression can be considered linear.

3 Answer to the question number 3:

From the given value in the question,we can write, $XY = [x_1, x_2, x_3, \ldots\ldots\ldots\ldots, x_n] \begin{bmatrix} y_1^T \\ y_1^T \\ . \\ . \\ . \\ y_n^T \end{bmatrix}$

Then after matrix multiplication, it becomes as follow:
$= [x_1 * y_1^T + x_2 * y_2^T + x_3 * y_3^T, \ldots\ldots\ldots\ldots, x_n * y_n^T]$
$X * Y = \sum x_i * y_i^T$ (Proved)

4 Problem 4:

For actual output y and predicted output $\hat{y}$, we can write residual error as follow :
$-(y - \hat{y})$
From the above the sum of the residual sum of square is:-
RSS $= \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$
$RSS = \sum (y_i - x_i * w)^2$
In matrix form it becomes:
RSS $= (y - x * w)^T (y - x * w)$
As we are differentiating w.r.t w, then our given equation from the question become as follow:
$-\frac{1}{2\sigma^2}$*RSS(w)
$= -\frac{1}{2\sigma^2} * (y - x * w)^T (y - x * w)$ (using the value of RSS)
$= -\frac{1}{2\sigma^2} * [y^T * y - w^T * x^T * y - y^T * x * w + w^T * x^T * x * w]$
$= -\frac{1}{2\sigma^2} * [y^T * y - 2 * w^T * x^T * y + w^T * x^T * x * w]$
By differentiating w.r.t w , we get:
$= -2x^T * y + 2 * x^T * x * w$
We can write it as $-2x^T * y + 2 * x^T * x * w = 0$
$w^* = (x^T x)^- 1 x^T y$
We can find optimal σ^2 by derivative of likelihood function w.r.t σ^2.
$\frac{d}{d(\sigma^2)} log L(w|x) = \frac{-1}{2}(\frac{-(RSS(w)}{\sigma^4}) + n/\sigma^2)$
If we consider
$RSS(w)/(\sigma^2)^2 - n/\sigma^2 = 0$
$\sigma^* = RSS(w)/n$

4.b

It is clear that just diving the RSS by number of sample we the optimal value of sigma.

5 Problem 5:

$Softmax(y_i) = exp(y_i)/\sum(exp(y_i)$

In binary case we have two possibilities, so the softmax looks like as follow –

$= exp(y_1)/[exp(y_1) + exp(y_2)]$

Dividing by $exp(y_1)$ we get,

$= 1/[1 + exp(y_2)/exp(y_1)]$

As we have two variables but one equation we can set y_2 to zero. Then, it becomes

$= 1/1 + exp(-y_1)$

6 Problem 6:

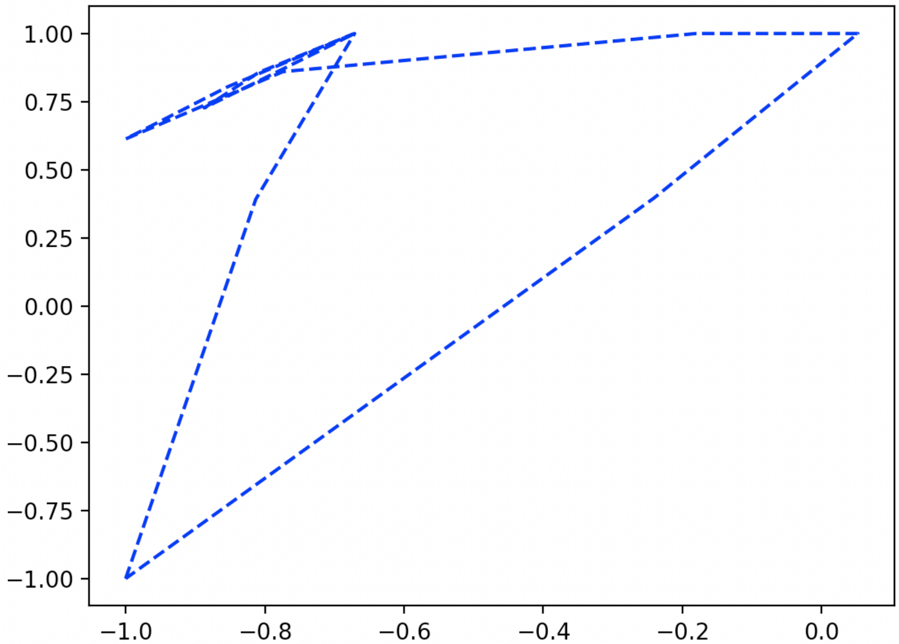

Figure 2: Question 6(a) image 1

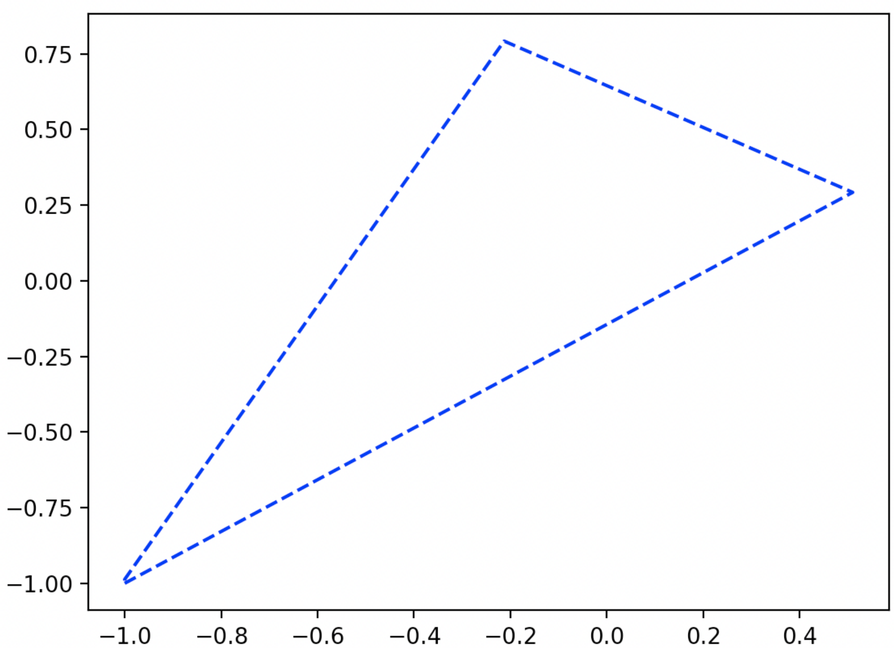

Figure 3: Question 6(a) image 2

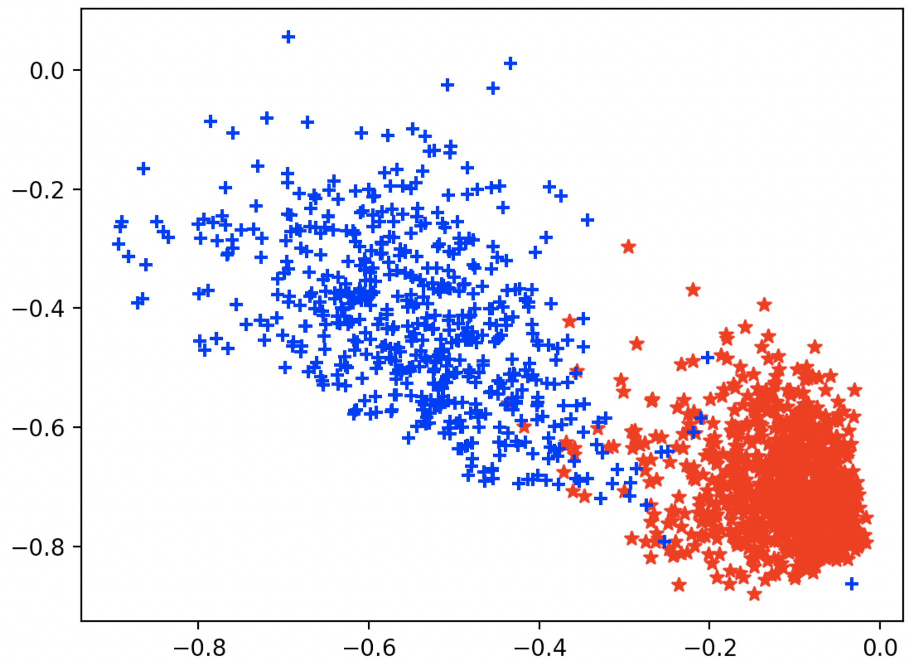

Figure 4: Question 6(b) features

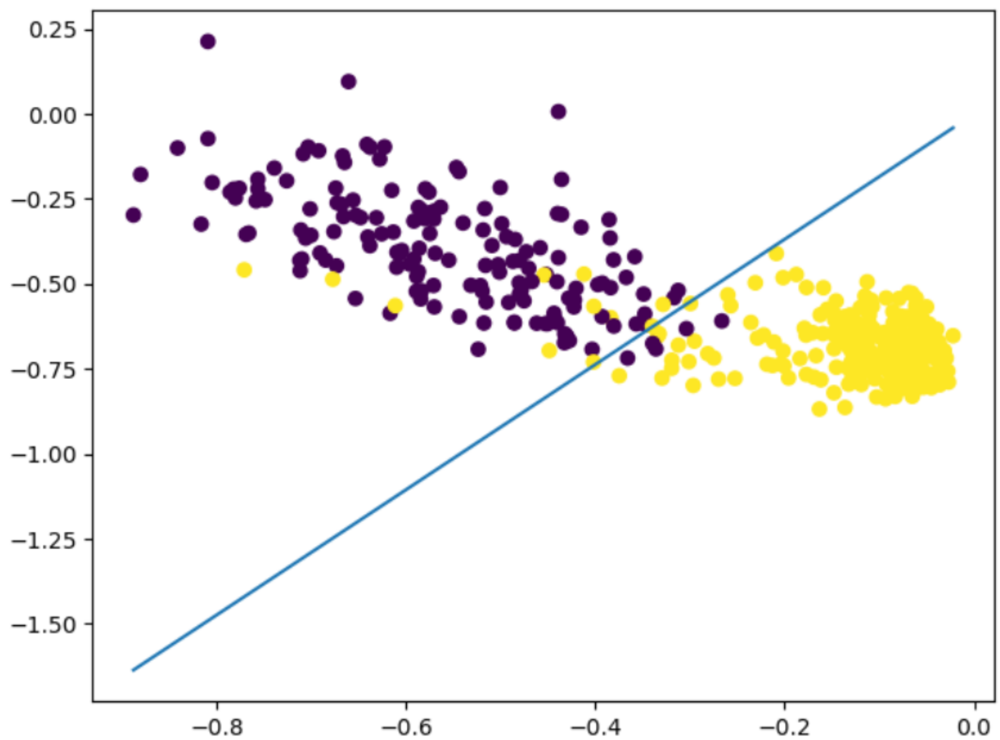

Figure 5: Question 6(d) decision boundary