Homework 5

Name: Yuan Li
ID: N19728558 netID yl6606
Email: foxerlee1@gmail.com

Part I: Written Exercises

1. (a)

Answer:

Because the decision boundary is 0.5, we can get the number of TP, FP, FN, TN should be:

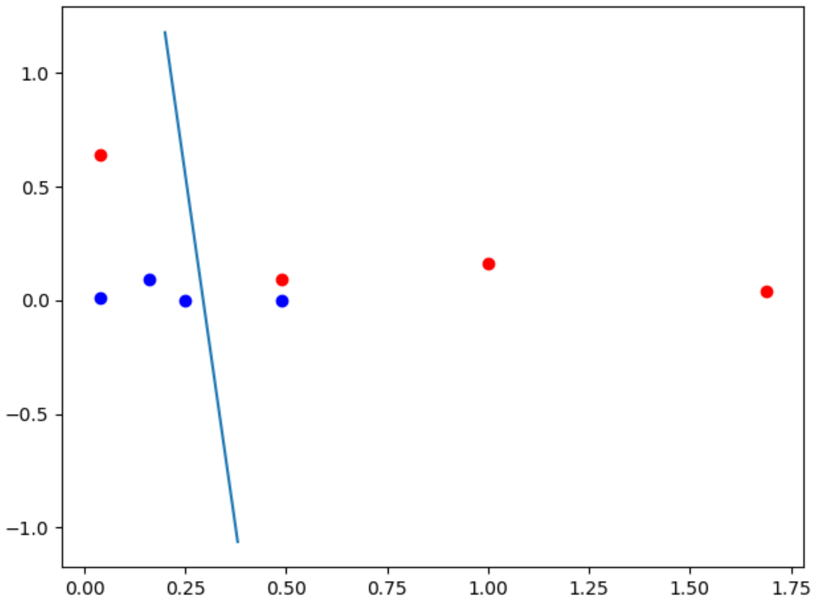
$$TP = 3$$
$$FP = 1$$
$$FN = 1$$
$$TN = 3$$

So the confusion matrix should be:

	pre=1	pre=0
true=1	3	1
true=0	1	3

1. (b)

Answer:



1. (c)

Answer:

$$FPR = \frac{FP}{TN + FP} = \frac{1}{1+3} = 0.25$$

1. (d)

Answer:

$$TPR = \frac{TP}{TP + FN} = \frac{3}{1+3} = 0.75$$

1. (e)

Answer:

$$accuracy = \frac{TP + TN}{TP + FP + FN + TN} = \frac{3+3}{1+3+1+3} = 0.75$$

1. (f)

Answer:

$$recall = \frac{TP}{TP + FP} = \frac{3}{1+3} = 0.75$$

1. (g)

Answer:

$$precision = \frac{TP}{TP + FN} = \frac{3}{1+3} = 0.75$$

1. (h)

Answer:

$$\begin{aligned}
cross(w) &= -(\sum_{i=1}^{N} y^{(i)} ln(h(x)) + (1 - y^{(i)}) ln(1 - h(x)))) \\
&= -(0 * ln0.389 + 1 * ln(1 - 0.389)) - (0 * ln0.042 + 1 * ln(1 - 0.042)) - (0 * ln0.613 + 1 * ln(1 - 0.613)) \\
&- (0 * ln0.167 + 1 * ln(1 - 0.167)) - (1 * ln0.572 + 0 * ln(1 - 0.572)) - (1 * ln0.526 + 0 * ln(1 - 0.526)) \\
&- (1 * ln0.393 + 0 * ln(1 - 0.393)) - (1 * ln0.638 + 0 * ln(1 - 0.638)) \\
&= 4.252
\end{aligned}$$

1. (i)

Answer:

We can calculate the error function for the new w'. Firstly, we need to get the new $h(x)$.

$$h(x) = [0.408, 0.022, 0.363, 0.112, 0.647, 0.643, 0.469, 0.765]$$

$$cross(w^{'}) = -(\sum_{i=1}^{N} y^{(i)} ln(h(x)) + (1-y^{(i)}) ln(1-h(x))))$$
$$= 3.015$$

Because $cross(w^{'}) < cross(w)$, $w^{'}$ is more likely to have generated the dataset given above.

1. (j)

Answer:

$$\begin{aligned} w_{new} &= w + \frac{\alpha}{N}\sum_{i=1}^{N}(y^{(i)} - h_\theta(x^{(i)}))x_i^{(i)} \\ &= [0.66, -2.24, -0.18]^T + \frac{0.1}{8}[0.66453595, 0.04747186, -0.41267559]^T \\ &= [0.668, -2.239, -0.185]^T \end{aligned}$$

1. (k)

Answer:

Considering the $(y^{(i)} - h_\theta(x^{(i)}))x_i^{(i)}$, this value for the data points near the decision boundary is big, for example, point $[1.0, 0.25, 0.0]$'s value is $[0.47502081, 0.1187552, 0.0]$ So they give big contribution.

1. (l)

Answer:

We can also consider the $(y^{(i)} - h_\theta(x^{(i)}))x_i^{(i)}$ for the data points which were correctly classified and far away from the decision boundary. For example, point $[1.0, 1.69, 0.04]$'s value is $[-0.041, -0.07, -0.001]$. So they give small contribution.

1. (m)

Answer:

Same as k and l, point $[1.0, 0.04, 0.64]$'s value is $[-0.611, -0.024, -0.391]$. So they give big contribution.

1. (n)

Answer:

$$l(w_{new}) = -(\sum_{i=1}^{N} y^{(i)} ln(h(x)) + (1-y^{(i)}) ln(1-h(x))))$$
$$= 4.245$$

1. (o)

Answer:

The cross-entropy went down after one iteration. This is my expected. Because it means the predict becomes more accurate, We are minimizing the error, our model is getting better, this is what our training expects.

2. (a)

Answer:

The learning rate is too large. So we move too far, which means overshoot.

2. (b)

Answer:

The learning rate is too small. So we move slowly, and each step is too small.

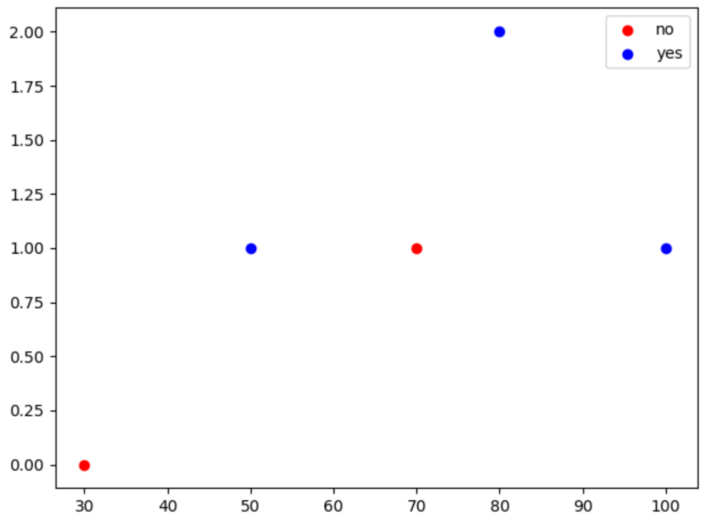
2. (c)

Answer:

This model is not suitable for this data.

3. (a)

Answer:



3. (b)

Answer:

I used the code from Part II to update the w. After 500 iterations, I got the result as:

$$w = \begin{bmatrix} -1.14371257 \\ 0.00838323 \\ 0.79041916 \end{bmatrix}$$

The predict labels are $[0, 1, 1, 1, 1]$, with only one error.

3. (c)

Answer:

i = 0. Because only $z^{(0)} < 0$, $-z^{(0)} > 0$, then $e^{-z^{(0)}}$ is the biggest. After we take the denominator, $\frac{1}{1+e^{-z^{(0)}}}$ should be the smallest.
So sample 0 is the least likely.

3. (d)

Answer:

The values y in part (b) don't change. Because the judgment in b and c is based on the values are positive or negative. If the original w multiplies a positive scalar, it will not change the values' positive or negative relationship. As a result, the result will not change.

The likelihoods in part (c) will change. They become:

$$P(y^{(i)} = 1|x^{(i)}, \alpha) = \frac{1}{1 + e^{-\alpha w^T x^{(i)}}}$$

So the change should be:

$$P^{'} = \frac{1}{1 + e^{-w^T x^{(i)}}} - \frac{1}{1 + e^{-\alpha w^T x^{(i)}}}$$