Homework 8

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

Part I: Written Exercises

1. (a)

Answer:

$$x_{local\_min} = -0.2792$$
$$x_{global\_min} = 1.9349$$

1. (b)

Answer:

At the start, x and $f(x)$ are:

$$x = -1$$
$$f(x) = 22$$

After 6 iterations, x and $f(x)$ are:

$$x_1 = -0.927$$
$$f(x)_1 = 17.0853$$
$$x_2 = -0.8651$$
$$f(x)_2 = 13.5381$$
$$x_3 = -0.8121$$
$$f(x)_3 = 10.9075$$
$$x_4 = -0.7659$$
$$f(x)_4 = 8.9130$$
$$x_5 = -0.7255$$
$$f(x)_5 = 7.3727$$
$$x_6 = -0.6897$$
$$f(x)_6 = 6.1646$$

For 1200 iterations, the last 6 values of x and $f(x)$ are all the same, which are:

$$x = -0.2792$$
$$f(x) = 0.4162$$

the value of x has converged. the gradient descent found a minimum. And it is a global minimum.

1. (c)

Answer:

When x starts with 3, x and $f(x)$ are:

$$x = 3$$
$$f(x) = 50$$

After 6 iterations, x and $f(x)$ are:

$$x_1 = 2.897$$
$$f(x)_1 = 40.3314$$
$$x_2 = 2.8119$$
$$f(x)_2 = 33.6675$$
$$x_3 = 2.7401$$
$$f(x)_3 = 28.8855$$
$$x_4 = 2.6785$$
$$f(x)_4 = 25.3431$$
$$x_5 = 2.6249$$
$$f(x)_5 = 22.6509$$
$$x_6 = 2.5779$$
$$f(x)_6 = 20.5609$$

For 1200 iterations, the last 6 values of x and $f(x)$ are all the same, which are:

$$x = 1.9349$$
$$f(x) = 9.9372$$

the value of x has converged. the gradient descent found a minimum. And it is a local minimum.

1. (d)

Answer:

When η changes to 0.01, At the start, x and $f(x)$ are:

$$x = -1$$
$$f(x) = 22$$

After 6 iterations, x and $f(x)$ are:

$$x_1 = -0.27$$
$$f(x)_1 = 0.4184$$
$$x_2 = -0.2746$$
$$f(x)_2 = 0.4167$$
$$x_3 = -0.2769$$
$$f(x)_3 = 0.4163$$
$$x_4 = -0.2781$$
$$f(x)_4 = 0.4162$$
$$x_5 = -0.2786$$
$$f(x)_5 = 0.4162$$
$$x_6 = -0.2789$$
$$f(x)_6 = 0.4162$$

For 1200 iterations, the last 6 values of x and $f(x)$ are all the same, which are:

$$x = -0.2792$$
$$f(x) = 0.4162$$

The value of x converged more quickly than $\eta = 0.001$.

1. (e)

Answer:

If I use python to calculate it, after several iterations, the value of $f(x)$ becomes too large to overflow.

2.

Answer:

For train sample one, The forward process shoule be:

$$z^{(2)} = w^{(1)} * x^{(1)} + b(1) = [2, 2]^T$$
$$a^{(2)} = [0.8807, 0.8807]^T$$
$$z^{(3)} = w^{(2)} * x^{(2)} + b(2) = 3.6423$$
$$a^{(3)} = 0.9744$$

Then the backward process should be:

$$\delta^{(3)} = -(y - a^{(3)})\Delta f(z^{(3)}) = -0.000634$$
$$\delta^{(2)} = w^{(2)} * \delta^{(3)}\Delta f(z^{(2)}) = [-6.6639 * 10^{-5} - 1.3327 * 10^{-4}]^T$$

So:

$$\Delta w^{(1)} = \Delta w^{(1)} + \delta^{(2)}a^{(1)} = \begin{bmatrix} -6.6639 * 10^{-5} & 0 \\ -1.3327 * 10^{-4} & 0 \end{bmatrix}$$
$$\Delta b^{(1)} = \Delta b^{(1)} + \delta^{(2)} = [-6.6639 * 10^{-5} - 1.3327 * 10^{-4}]^T$$
$$\Delta w^{(2)} = \Delta w^{(2)} + \delta^{(3)}a^{(2)} = [-5.5904 * 10^{-4} - 5.5904 * 10^{-4}]$$
$$\Delta b^{(2)} = \Delta b^{(2)} + \delta^{(3)} = -6.6347 * 10^{-4}$$

For train sample two, The forward process shoule be:

$$z^{(2)} = w^{(1)} * x^{(1)} + b(1) = [33]^T$$
$$a^{(2)} = [0.9525, 0.9525]^T$$
$$z^{(3)} = w^{(2)} * x^{(2)} + b(2) = 3.8577$$
$$a^{(3)} = 0.9793$$

Then the backward process should be:

$$\delta^{(3)} = -(y - a^{(3)})\Delta f(z^{(3)}) = 0.01983$$
$$\delta^{(2)} = w^{(2)} * \delta^{(3)}\Delta f(z^{(2)}) = [-8.9598 * 10^{-4}, -1.7919 * 10^{-3}]^T$$

So:

$$\Delta w^{(1)} = \Delta w^{(1)} + \delta^{(2)}a^{(1)} = \begin{bmatrix} -6.6639 * 10^{-5} & -8.9598 * 10^{-4} \\ -1.3327 * 10^{-4} & -1.7919 * 10^{-3} \end{bmatrix}$$
$$\Delta b^{(1)} = \Delta b^{(1)} + \delta^{(2)} = [-8.9598 * 10^{-4}, -1.7919 * 10^{-3}]^T$$
$$\Delta w^{(2)} = \Delta w^{(2)} + \delta^{(3)}a^{(2)} = [0.018330.01833]$$
$$\Delta b^{(2)} = \Delta b^{(2)} + \delta^{(3)} = 0.01919$$

Then we can get new w and b.

$$w^{(1)} = w^{(1)} - \alpha/N * \Delta w^{(1)} = \begin{bmatrix} 1.000007 & 1.999 \\ 3.000013 & 3.999 \end{bmatrix}$$
$$b^{(1)} = b^{(1)} - \alpha/N * \Delta b^{(1)} = [0.9999 - 1.0016]^T$$
$$w^{(2)} = w^{(2)} - \alpha/N * \Delta w^{(2)} = [0.99811.998]$$
$$b^{(2)} = b^{(2)} - \alpha/N * \Delta b^{(2)} = 0.9980$$

3. (a)

Answer:

NeuralNetRK can output values that are negative numbers.

3. (b)

Answer:

NeuralNetCK.

3. (c)

Answer:

NeuralNetCK would be appropriate. Because this problem is a classification problem and it has three categories, more than two.

3. (d)

Answer:

NeuralNetZeroOne would be appropriate. We can consider this problem has two classification problems. First one is whether or not the document is about politics, and the second one is whether it is written in a formal or informal style. So **NeuralNetZeroOne** would be better.

4. (a)

Answer:

Because forest tree is insensitive to values, which means if we change the numerical values of nominal attributes, the results would not change. However, neural net is sensitive to values. For example, if we set almond(1000) or other big value, we may get a different result.

4. (b) i.

Answer:

	z1	z2	z3	z4	z5	z6	label
$x^{(1)}$	0	0	0	1	1	0	0
$x^{(2)}$	0	0	1	0	0	1	0

4. (b) ii.

Answer:

For ordinal attributes like 1, 2 and 3, they will include the order relationship, like 1 is smaller than 2, and 2 is smaller than 3. But if we use one-hot-encoding, they will not include the order relationship.

4. (b) iii.

Answer:

For nominal attributes with only two values, we can set 1 means "has x1, not x2", and 0 means "has x2, not x1". This effect is the same as one-hot-encoding. However, if attributes with more than two values, then it will become 1, 2, 3, and so on. We have discussed about the disadvantage before.

5.

Answer:

The size of the result should be $(W - F + 1) * (W - F + 1) = 2 * 2$. And the result is as follow:

275	275
248	185

6. *Answer:*

We should add padding with $P = 1$. The size of the activation map should be:

$$\begin{aligned} size &= (W + 2P - F + 1) * (W + 2P - F + 1) \\ &= (4 + 2 - 3 + 1) * (4 + 2 - 3 + 1) \\ &= 4 * 4 \end{aligned}$$

-288	-367	-244	-121
164	275	275	176
183	248	185	97
198	225	114	42

7. *Answer:*

The size of the activation map should be:

$$\begin{aligned} size &= ((W + 2P - F)/S + 1) * ((W + 2P - F)/S + 1) \\ &= ((4 + 2 - 2)/2 + 1) * ((4 + 2 - 2)/2 + 1) \\ &= 3 * 3 \end{aligned}$$