COMP4434 Homework 3

ZHANG Caiqi 18085481d

April 17, 2021

1 Question 1

- the graphs are shown in Fig. 1.
- Without regularization: 2.6402468133652626. In this case, the model is overfitting. We will continue the following analysis using this one as the baseline.
- L1 and $\lambda = 1$: 2.605842524474064. Compared with the baseline, the RMSE decrease and the line becomes flatter. However, compared with the result of L2 with $\lambda = 1$, the RMSE still can be improved. Therefore, in this case, the model is slightly underfitteting or, to some extent, almost fitted.
- L1 and $\lambda = 100$: 3.4324281069662836. The RMSE is larger than the baseline and the line in the graph is too flat, which means it cannot match the dataset well. In this case, the model is underfitting.
- L2 and $\lambda = 1$: 2.5823732857847608. This is the best result among the four experiments. In this case, the model is almost fitted. If we also want to make sure whether this is slightly underfitting or overfitting, we need to conduct extra experiments.
- L2 and $\lambda = 100$: 2.7301727313593505. In this case, the model is underfitting. Similar reason as L1 and $\lambda = 100$.

2 Question 2

2.1 a)

$x^{(1)} = [1.268\ 0.994]^T$ and $\theta^{(3)} = [0.271\ 0.694]^T$. $x^{(1)} \cdot \theta^{(3)} = 1.268 \times 0.271 + 0.994 \times 0.694 = 1.033$.

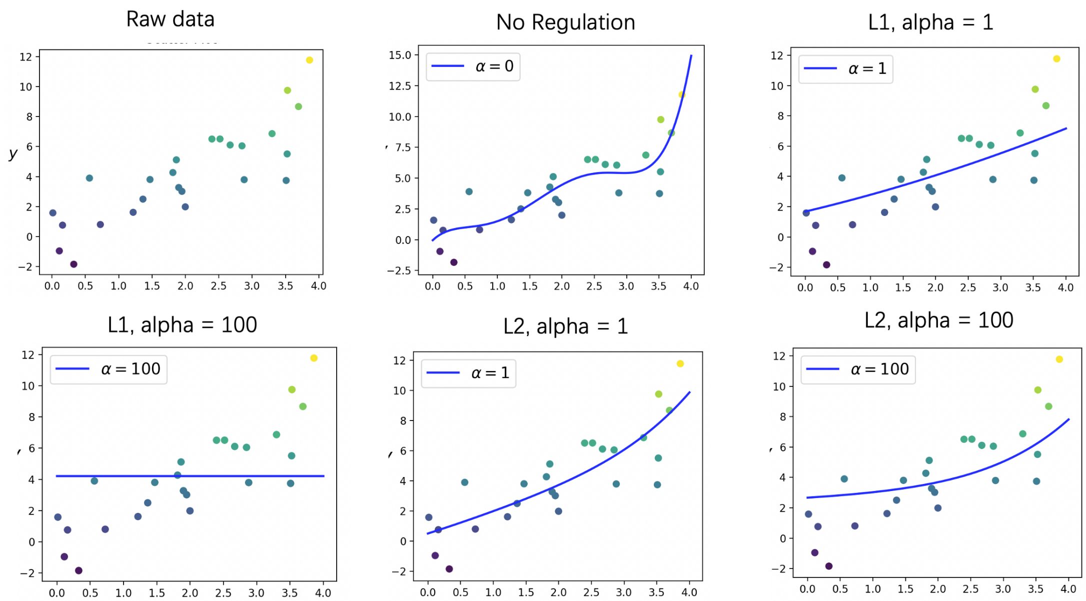

Figure 1: Experiment results

2.2 b)

- For $x_1^{(1)}$, the gradient is: $(1/9) \times ((0.77 \times 0.19 + 0.48 \times 0.62 - 5) \times 0.19 + (0.77 \times 0.68 + 0.48 \times 0.78 - 4) \times 0.68 + (0.77 \times 0.36 + 0.48 \times 0.92 - 3) \times 0.36 + 0.02 \times 0.77) = -0.420$. Therefore, $x_1^{(1)} = 0.77 - \alpha * -0.420 = 0.770$.

- For $\theta_1^{(2)}$, the gradient is: $(1/9) \times ((0.68 \times 0.77 + 0.78 \times 0.48 - 4) \times 0.77 + (0.68 \times 0.31 + 0.78 \times 0.51 - 3) \times 0.31 + 0.02 \times 0.68) = -0.346$. Therefore, $\theta_1^{(2)} = 0.68 - 0.0002 \times (-0.0346) = 0.680$.

2.3 c)

$x^{(2)} = [0.432\ 0.443]^T$ and $\theta^{(2)} = [0.683\ 0.783]^T$, therefore the rate is $0.432 \times 0.683 + 0.443 \times 0.783 = 0.642$. Deatils can be found in the code file.

3 Question 3

3.1 a)

$$
\begin{aligned}
z_4 &= sigmoid(a_4) \\
&= sigmoid(z_1 \cdot w_1^4 + z_3 \cdot w_3^4 +_2 \cdot w_2^4) \\
&= sigmoid(sigmoid(a_1) \cdot w_1^4 + relu(a_3) \cdot w_3^4 + sigmoid(a_2) \cdot w_2^4) \\
&= sigmoid(sigmoid(a_1) \cdot w_1^4 + relu(z_1 \cdot w_1^3 + z_2 \cdot w_2^3) \cdot w_3^4 + sigmoid(a_2) \cdot w_2^4) \\
&= sigmoid(sigmoid(a_1) \cdot w_1^4 + relu(sigmoid(a_1) \cdot w_1^3 + sigmoid(a_2) \cdot w_2^3) \cdot w_3^4 + sigmoid(a_2) \cdot w_2^4) \\
&= sigmoid(sigmoid(z_0 \cdot w_0^1) \cdot w_1^4 + relu(sigmoid(z_0 \cdot w_0^1) \cdot w_1^3 + sigmoid(z_0 \cdot w_0^2) \cdot w_2^3) \cdot w_3^4 + \\
&\quad sigmoid(z_0 \cdot w_0^2) \cdot w_2^4) \\
&= sigmoid(sigmoid(a_0 \cdot w_0^1) \cdot w_1^4 + relu(sigmoid(a_0 \cdot w_0^1) \cdot w_1^3 + sigmoid(a_0 \cdot w_0^2) \cdot w_2^3) \cdot w_3^4 + \\
&\quad sigmoid(a_0 \cdot w_0^2) \cdot w_2^4)
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
z_4 &= F(a_0) \\
&= sigmoid(sigmoid(a_0 \cdot w_0^1) \cdot w_1^4 + relu(sigmoid(a_0 \cdot w_0^1) \cdot w_1^3 + sigmoid(a_0 \cdot w_0^2) \cdot w_2^3) \cdot w_3^4 + \\
&\quad sigmoid(a_0 \cdot w_0^2) \cdot w_2^4)
\end{aligned}
$$

3.2 b)

We denote $sigmoid(a)$ as $s(a)$.

$$
\begin{aligned}
\frac{\partial J}{\partial w_3^4} &= \frac{\partial \frac{1}{2}(z_4 - y)^2}{\partial w_3^4} \\
&= \frac{\partial \frac{1}{2}(z_4 - y)^2}{\partial z_4} \cdot \frac{\partial z_4}{\partial w_3^4} \\
&= (z_4 - y) \cdot \frac{\partial s(a_4)}{\partial w_3^4} \\
&= (z_4 - y) \cdot \frac{\partial s(a_4)}{\partial a_4} \cdot \frac{\partial a_4}{\partial w_3^4} \\
&= (z_4 - y) \cdot s(a_4) \cdot (1 - s(a_4)) \cdot \frac{\partial a_4}{\partial w_3^4} \\
&= (z_4 - y) \cdot s(a_4) \cdot (1 - s(a_4)) \cdot \frac{\partial (z_1 \cdot w_1^4 + z_3 \cdot w_3^4 +_2 \cdot w_2^4)}{\partial w_3^4} \\
&= (z_4 - y) \cdot s(a_4) \cdot (1 - s(a_4)) \cdot z_3 \\
&= (z_4 - y) \cdot z_4 \cdot (1 - z_4) \cdot z_3
\end{aligned}
$$

If we want to only use a_0 to represent the result, we may plug in the followings:

$$
\begin{aligned}
z_4 &= sigmoid(sigmoid(a_0 \cdot w_0^1) \cdot w_1^4 + relu(sigmoid(a_0 \cdot w_0^1) \cdot w_1^3 + sigmoid(a_0 \cdot w_0^2) \cdot w_2^3) \cdot w_3^4 + \\
&\quad sigmoid(a_0 \cdot w_0^2) \cdot w_2^4) \\
z_3 &= relu(sigmoid(a_0 \cdot w_0^1) \cdot w_1^3 + sigmoid(a_0 \cdot w_0^2) \cdot w_2^3) \\
y &= 0.1 \\
a_0 &= 1
\end{aligned}
$$