Machine Learning Homework

Jia-Yong Shen
7107053114

I. Description

A. Step

step 1:
(1) Prepare the Dataset
(2) Pre-process the Dataset

step 2:
Forward Propagation :
→ start at the input layer
→ forward propagate the data through
→ the neural network
→ generate an output

step 3:
Calculate Training Cost :
→ Calculate cost

step 4:
Backward Propagation :
→ backpropagate the error
→ find each weight
→ update the model
→ minimize the cost

II. Problem 1

A. Problem Description

Derive the forward and backward schemes for the Regression Problem, where the cost is $\jmath(W) = \sum_i \left\| \hat{y}^{(i)} - y^{(i)} \right\|^2$

B. Forward

1) Hidden Layer: One unit in the input layer : Net input

$$Z_1^{(h)} = \begin{bmatrix} a_1^{(in)} & a_2^{(in)} & \cdots & a_m^{(in)} \end{bmatrix}^T \begin{bmatrix} W_{0,1}^{(h)} & W_{1,1}^{(h)} & \cdots & W_{m,1}^{(h)} \end{bmatrix} \tag{1}$$

One unit in the input layer : After the activation function

$$a_1^{(h)} = \phi\left(Z_1^{(h)}\right) \tag{2}$$

All units in the input layer : Net input

$$\begin{aligned} \underset{z}{\rightarrow}^{(h)} &= \begin{bmatrix} z_1^{(h)} & z_2^{(h)} & \cdots & z_n^{(h)} \end{bmatrix}^T \\ &= \begin{bmatrix} \underset{w_1}{\rightarrow}^{(h)} & \underset{w_2}{\rightarrow}^{(h)} & \cdots & \underset{w_n}{\rightarrow}^{(h)} \end{bmatrix}^T \underset{a}{\rightarrow}^{(in)} \\ &= \left(W^{(h)}\right)^T \underset{a}{\rightarrow}^{(in)} \end{aligned} \tag{3}$$

All units in the input layer : After the activation function

$$\begin{aligned} \underset{a}{\rightarrow}^{(h)} &= \phi\left(\left(W^{(h)}\right)^T \underset{a}{\rightarrow}^{(in)}\right) \\ &= \phi\left(\underset{z}{\rightarrow}^{(h)}\right) \end{aligned} \tag{4}$$

We have:

$$\begin{aligned} Z^{(h)} &= \left(W^{(h)}\right)^T A^{(in)} \\ A^{(h)} &= \phi\left(Z^{(h)}\right) \end{aligned} \tag{5}$$

2) Output Layer: All units in the hidden layer : Net input

$$\begin{aligned} \underset{z}{\rightarrow}^{(out)} &= \begin{bmatrix} \underset{w_1}{\rightarrow}^{(out)} & \underset{w_2}{\rightarrow}^{(out)} & \cdots & \underset{w_n}{\rightarrow}^{(out)} \end{bmatrix}^T \underset{a}{\rightarrow}^{(h)} \\ &= \left(W^{(out)}\right)^T \underset{a}{\rightarrow}^{(h)} \\ &= \left(W^{(out)}\right)^T A^{(h)} \end{aligned} \tag{6}$$

All units in the hidden layer : Output

$$\begin{aligned} Z^{(out)} &= \left(W^{(out)}\right)^T A^{(h)} \\ A^{(out)} &= Z^{(out)} \end{aligned} \tag{7}$$

C. Backward

1) Output Layer:

$$\begin{aligned} \frac{\partial \jmath(W)}{\partial W^{(out)}} &= \frac{\partial\left(\left(\hat{y}^{(i)} - y\right)\left(\hat{y}^{(i)} - y\right)^T\right)}{\partial\left(\hat{y}^{(i)} - y\right)} \cdot \frac{\partial\left(\hat{y}^{(i)} - y\right)}{\partial W^{(out)}} \\ &= 2 \cdot (\hat{y} - y) \cdot \frac{\partial\left(\hat{y}^{(i)} - y\right)}{\partial W^{(out)}} \\ &= 2 \cdot (\hat{y} - y) \cdot \frac{\partial\left(W^{(out)} A^{(h)} - y\right)}{\partial W^{(out)}} \\ &= 2 \cdot (\hat{y} - y) \cdot A^{(h)} \\ &= 2 \cdot \delta^{(out)} \cdot A^{(h)} \end{aligned} \tag{8}$$

We have :

$$\frac{\partial \jmath(W)}{\partial W^{(out)}} = 2 \cdot \delta^{(out)} \cdot A^{(h)} \tag{9}$$

Notes:

$$\jmath(W) = \sum_i \left\| \hat{y}^{(i)} - y^{(i)} \right\|^2 = (\hat{y} - y)(\hat{y} - y)^T \tag{10}$$

$$\delta^{(out)} = \hat{y} - y \tag{11}$$

$$\hat{y} = A^{(out)} = Z^{(out)} = W^{(out)} A^{(h)} \tag{12}$$

2) Hidden Layer:

$$\begin{aligned}
\frac{\partial \jmath(W)}{\partial W^{(h)}} &= \frac{\partial \left(\left(\hat{y}^{(i)} - y\right)\left(\hat{y}^{(i)} - y\right)^T \right)}{\partial \left(\hat{y}^{(i)} - y\right)} \cdot \frac{\partial \left(\hat{y}^{(i)} - y\right)}{\partial W^{(h)}} \\
&= 2 \cdot (\hat{y} - y) \cdot \frac{\partial \left(\hat{y}^{(i)} - y\right)}{\partial W^{(h)}} \\
&= 2 \cdot (\hat{y} - y) \cdot \frac{\partial \left(W^{(out)} A^{(h)} - y\right)}{\partial W^{(h)}} \\
&= 2 \cdot (\hat{y} - y) \cdot \frac{\partial \left(W^{(out)} A^{(h)} - y\right)}{\partial \left(W^{(out)} A^{(h)}\right)} \cdot \frac{\partial \left(W^{(out)} A^{(h)}\right)}{\partial W^{(h)}} \\
&= 2 \cdot (\hat{y} - y) \cdot \frac{\partial \left(W^{(out)} A^{(h)}\right)}{\partial W^{(h)}} \\
&= 2 \cdot (\hat{y} - y) \cdot \frac{\partial \left(W^{(out)} A^{(h)}\right)}{\partial A^{(h)}} \cdot \frac{\partial A^{(h)}}{\partial W^{(h)}} \\
&= 2 \cdot (\hat{y} - y) \cdot W^{(out)} \cdot \frac{\partial A^{(h)}}{\partial W^{(h)}} \\
&= 2 \cdot (\hat{y} - y) \cdot W^{(out)} \cdot \frac{\partial \left(\phi\left(Z^{(h)}\right)\right)}{\partial Z^{(h)}} \cdot \frac{\partial Z^{(h)}}{\partial W^{(h)}} \\
&= 2 \cdot (\hat{y} - y) \cdot W^{(out)} \cdot \frac{\partial \left(\phi\left(Z^{(h)}\right)\right)}{\partial Z^{(h)}} \cdot \frac{\partial W^{(h)} A^{(in)}}{\partial W^{(h)}} \\
&= 2 \cdot (\hat{y} - y) \cdot W^{(out)} \cdot A^{(in)} \cdot \frac{\partial \left(\phi\left(Z^{(h)}\right)\right)}{\partial Z^{(h)}}
\end{aligned} \tag{13}$$

We have :

$$\frac{\partial \jmath(W)}{\partial W^{(h)}} = 2 \cdot (\hat{y} - y) \cdot W^{(out)} \cdot A^{(in)} \cdot \frac{\partial \left(\phi\left(Z^{(h)}\right)\right)}{\partial Z^{(h)}} \tag{14}$$

Notes:

$$A^{(h)} = \phi\left(Z^{(h)}\right) \tag{15}$$

$$Z^{(h)} = W^{(h)} A^{(in)} \tag{16}$$

III. Problem 2

A. Problem Description

Prepare the house data set, and then do the preprocessing to the data.

B. Prepare the house dataset

Download the housing dataset.

C. Preprocessing

Normalization X and normalization y.

IV. Problem 3

A. Problem Description

Implement the 3-layer MLP (including the cost, forward and backward schemes) for the regression problem.

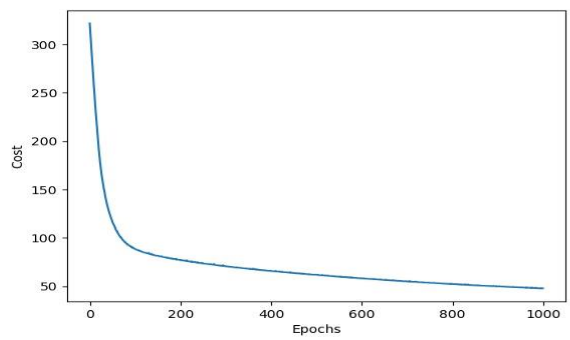

Fig. 1. Cost