Neural Network Back Propagation 公式推导

111

2017年3月20日

1 Notions

L: neural network layer number

s_l: 第l层神经元个数

$z_i^{(l)}$: 第l层第i个神经元的输入

$a_i^{(l)}$: 第l层第i个神经元的输出

$W_{ij}^{(l)}$: 第$l-1$层第j个神经元到第l层第i个神经元的权重，维度：$s_l \times s_{l-1}$

第二层各神经元的计算方法如下：

$$
\begin{aligned}
a_1^{(2)} &= f(W_{11}^{(2)}x_1 + W_{12}^{(2)}x_2 + W_{13}^{(2)}x_3 + b_1^{(2)}) \\
a_2^{(2)} &= f(W_{21}^{(2)}x_1 + W_{22}^{(2)}x_2 + W_{23}^{(2)}x_3 + b_2^{(2)}) \\
a_3^{(2)} &= f(W_{31}^{(2)}x_1 + W_{32}^{(2)}x_2 + W_{33}^{(2)}x_3 + b_3^{(2)}) \\
a_4^{(2)} &= f(W_{41}^{(2)}x_1 + W_{42}^{(2)}x_2 + W_{43}^{(2)}x_3 + b_4^{(2)})
\end{aligned}
$$

其中，$f()$为activation function. 向量表示为：

$$
\begin{aligned}
z^{(2)} &= W^{(2)}x + b^{(2)} \\
a^{(2)} &= f(z^{(2)})
\end{aligned}
$$

矩阵W为：

$$
W_{4\times 3} = \begin{bmatrix}
W_{11}^{(2)} & W_{12}^{(2)} & W_{13}^{(2)} \\
W_{21}^{(2)} & W_{22}^{(2)} & W_{23}^{(2)} \\
W_{31}^{(2)} & W_{32}^{(2)} & W_{33}^{(2)} \\
W_{41}^{(2)} & W_{42}^{(2)} & W_{43}^{(2)}
\end{bmatrix}
$$

feedward的代码见[1]

2 Activation function

激活函数种类很多。

[1] network.py中的函数 feedforward(self, a)

sigmoid funcrion[2]。

$$\sigma(z) = \frac{1}{1+\exp(-z)} \tag{1}$$

对应的微分方程[3]:

$$\sigma'(z) = \sigma(z) * (1 - \sigma(z)) \tag{2}$$

3 Cost function

1个样本的损失函数为:

$$J(W,b;x,y) = \frac{1}{2}\sum_i (a_i^{(L)} - y_i)^2 \tag{3}$$

其中,i为输出层(output layer)神经元个数。

4 Error of neuron

The error of neuron i in layer l is defined as follow:

$$\delta_i^{(l)} = \frac{\partial}{\partial z_i^{(l)}} J(W,b;x,y) \tag{4}$$

又由于 $a_i^{(l)} = f(z_i^{(l)})$, 公式 (4) 可以转换成:

$$\delta_i^{(l)} = \frac{\partial}{\partial a_i^{(l)}} J(W,b;x,y) f'(z_i^{(l)}) \tag{5}$$

根据公式 (3)可知,输出层(output layer)神经元偏差(error of neuron)为:

$$\frac{\partial}{\partial a_i^{(L)}} J(W,b;x,y) = (a_i^{(L)} - y_i) \tag{6}$$

向量表示输出层(output layer)偏差(error of neuron)为:

$$\delta^{(L)} = (a^L - y) \bigodot f'(z^L) \tag{7}$$

其它层偏差的向量表示为[4]:

$$\delta^l = (W^{(l+1)})^T \delta^{(l+1)} \bigodot f'(z^{(l)}) \tag{8}$$

[2]network.py中的函数 sigmoid(z)
[3]network.py中的函数 sigmoid_prime(z)
[4]此处证明见section (7)

5 Error of w and b

$$\begin{aligned}
\frac{\partial}{\partial W_{ij}^{(l)}} J(W,b;x,y) &= \frac{\partial J(W,b;x,y)}{\partial z_i^{(l)}} \frac{\partial z_i^{(l)}}{\partial W_{ij}^{(l)}} \\
&= \frac{\partial J(W,b;x,y)}{\partial z_i^{(l)}} a_j^{(l-1)} \\
&= \delta_i^{(l)} a_j^{(l-1)}
\end{aligned} \tag{9}$$

向量表示为：

$$\frac{\partial J}{\partial W^{(l)}} = \delta^{(l)} (a^{(l-1)})^T \tag{10}$$

$$\begin{aligned}
\frac{\partial}{\partial b_i^{(l)}} J(W,b;x,y) &= \frac{\partial J(W,b;x,y)}{\partial z_i^{(l)}} \frac{\partial z_i^{(l)}}{\partial b_i^{(l)}} \\
&= \frac{\partial J(W,b;x,y)}{\partial z_i^{(l)}} \\
&= \delta_i^{(l)}
\end{aligned} \tag{11}$$

向量表示为：

$$\frac{\partial J}{\partial b^{(l)}} = \delta^{(l)} \tag{12}$$

6 Update parameters

K个样本对应的偏差加起来：

$$\begin{aligned}
\triangle W^{(l)} &:= \triangle W^{(l)} + \frac{\partial J}{\partial W^{(l)}} \\
\triangle b^{(l)} &:= \triangle b^{(l)} + \frac{\partial J}{\partial b^{(l)}}
\end{aligned}$$

更新一次$W^{(l)}$和$b^{(l)}$:

$$\begin{aligned}
W^{(l)} &= W^{(l)} - \alpha \left[\left(\frac{1}{K} \triangle W^{(l)} \right) + \lambda W^{(l)} \right] \\
b^{(l)} &= b^{(l)} - \alpha \left[\left(\frac{1}{K} \triangle b^{(l)} \right) \right]
\end{aligned}$$

7 证明 δ^l 与 δ^{l+1} 之间的关系

由公式 (8) 可知输出层神经元的偏差，现在计算导数第二次神经元偏差：

$$\begin{aligned}
\delta_i^{(L-1)} &= \frac{\partial}{\partial z_i^{(L-1)}} J(W,b;x,y) \\
&= \frac{\partial}{\partial z_i^{(L-1)}} \frac{1}{2} \sum_{k=1}^{s_k} (a_k^{(L)} - y_k)^2 \\
&= \frac{1}{2} \sum_{k=1}^{s_k} \frac{\partial}{\partial z_i^{(L-1)}} (a_k^{(L)} - y_k)^2 \\
&= \sum_{k=1}^{s_k} (a_k^{(L)} - y_k) \frac{\partial a_k^{(L)}}{\partial z_i^{(L-1)}} \\
&= \sum_{k=1}^{s_k} (a_k^{(L)} - y_k) \frac{\partial a_k^{(L)}}{\partial z_k^{(L)}} \frac{\partial z_k^{(L)}}{\partial z_i^{(L-1)}} \\
&= \sum_{k=1}^{s_k} (a_k^{(L)} - y_k) f'(z_k^{(L)}) \frac{\partial z_k^{(L)}}{\partial z_i^{(L-1)}} \\
&= \sum_{k=1}^{s_k} \delta_k^{(L)} W_{ki}^{(L)} f'(z_i^{(L-1)})
\end{aligned} \tag{13}$$

上式中，

$$\begin{aligned}
z_k^{(L)} &= \sum_{j=1}^{s_{(L-1)}} W_{kj}^{(L)} a_j^{(L-1)} + b_k^{(L)} \\
&= \sum_{j=1}^{s_{(L-1)}} W_{kj}^{(L)} f(z_j^{(L-1)}) + b_k^{(L)}
\end{aligned}$$

所以，

$$\begin{aligned}
\frac{\partial z_k^{(L)}}{\partial z_i^{(L-1)}} &= \frac{\partial}{\partial z_i^{(L-1)}} \left[\sum_{j=1}^{s_{(L-1)}} W_{kj}^{(L)} f(z_j^{(L-1)}) + b_k^{(L)} \right] \\
&= \frac{\partial}{\partial z_i^{(L-1)}} \left[W_{ki}^{(L)} f(z_i^{(L-1)}) + b_k^{(L)} \right] \\
&= W_{ki}^{(L)} f'(z_i^{(L-1)})
\end{aligned}$$

由公式 (13) 推广到一般形式：

$$\delta_i^{(l)} = \sum_{k=1}^{s_{l+1}} \delta_k^{(l+1)} W_{ki}^{(l+1)} f'(z_i^{(l)}) \tag{14}$$

向量形式表示为：

$$\delta^{(l)} = (W^{(l+1)})^T \delta^{(l+1)} \odot f'(z^{(l)}) \tag{15}$$

故，公式 (8) 得证。