EMS702P Statistical Thinking and Applied Machine Learning

Week 9 – PBL Solutions

Yunpeng Zhu

Solution 1:

(a):

According to the initial settings of $\bar{w}_1 = 0.3$ and $\bar{w}_2 = 0.1$, the initial outputs of the neural network is

$$\begin{cases} \bar{z}_1(1) = \bar{w}_1 x(1) = 0.3 \times 1 = 0.3 \\ \bar{z}_1(2) = \bar{w}_1 x(2) = 0.3 \times 1.5 = 0.45 \\ \bar{z}_1(3) = \bar{w}_1 x(3) = 0.3 \times 2 = 0.6 \end{cases} \tag{1}$$

$$\begin{cases} f\left[\bar{z}_1(1)\right] = \dfrac{1}{1+\exp\left[-\bar{z}_1(1)\right]} = \dfrac{1}{1+\exp(-0.3)} = 0.574 \\ f\left[\bar{z}_1(2)\right] = \dfrac{1}{1+\exp\left[-\bar{z}_1(2)\right]} = \dfrac{1}{1+\exp(-0.45)} = 0.611 \\ f\left[\bar{z}_1(3)\right] = \dfrac{1}{1+\exp\left[-\bar{z}_1(3)\right]} = \dfrac{1}{1+\exp(-0.6)} = 0.646 \end{cases} \tag{2}$$

$$\begin{cases} \bar{y}(1) = \bar{z}_2(1) = \bar{w}_2 f\left[\bar{z}_1(1)\right] = 0.1 \times 0.574 = 0.0574 \\ \bar{y}(2) = \bar{z}_2(2) = \bar{w}_2 f\left[\bar{z}_1(2)\right] = 0.1 \times 0.611 = 0.0611 \\ \bar{y}(3) = \bar{z}_2(3) = \bar{w}_2 f\left[\bar{z}_1(3)\right] = 0.1 \times 0.646 = 0.0646 \end{cases} \tag{3}$$

The cost function of the neural network is

$$\begin{aligned} J &= \frac{J_1 + J_2 + J_3}{3} \\ &= \frac{1}{3}\left\{\frac{1}{2}\left[\bar{y}(1) - y(1)\right]^2 + \frac{1}{2}\left[\bar{y}(2) - y(2)\right]^2 + \frac{1}{2}\left[\bar{y}(3) - y(3)\right]^2\right\} \end{aligned} \tag{4}$$

Now calculate the gradients separately:

(b):

The gradient of J with respect to the weight w_2 is obtained as

$$\frac{\partial J}{\partial w_2} = \frac{1}{3}\left[\frac{\partial J_1}{\partial w_2} + \frac{\partial J_2}{\partial w_2} + \frac{\partial J_3}{\partial w_2}\right] \tag{5}$$

in which the gradients of $J_t, t = 1, 2, 3$ with respect to w_2 are obtained as

$$\frac{\partial J_t}{\partial w_2} = \left[\bar{y}(t) - y(t)\right] \times \left.\frac{\partial y}{\partial w_2}\right|_{y=y(t)}, k = 1,2,3 \tag{6}$$

where

$$\frac{\partial y}{\partial w_2} = \frac{\partial\left[f(z_1)w_2\right]}{\partial w_2} = f(z_1) \tag{7}$$

Therefore,

$$\begin{cases} \dfrac{\partial J_1}{\partial w_2} = \left[\bar{y}(1) - y(1)\right] f\left[\bar{z}_1(1)\right] = (0.0574 - 0.3) \times 0.574 = -0.139 \\ \dfrac{\partial J_2}{\partial w_2} = \left[\bar{y}(2) - y(2)\right] f\left[\bar{z}_1(2)\right] = (0.0611 - 0.6) \times 0.611 = -0.329 \\ \dfrac{\partial J_3}{\partial w_2} = \left[\bar{y}(3) - y(3)\right] f\left[\bar{z}_1(3)\right] = (0.0646 - 0.8) \times 0.646 = -0.475 \end{cases} \tag{8}$$

Substituting (8) into gradient (5), yields

$$\frac{\partial J}{\partial w_2} = \frac{1}{3} \times (-0.139 - 0.329 - 0.475) = -0.314 \tag{9}$$

(c):

The gradient of J with respect to the weight w_1 is obtained as

$$\frac{\partial J}{\partial w_1} = \frac{1}{3}\left[\frac{\partial J_1}{\partial w_1} + \frac{\partial J_2}{\partial w_1} + \frac{\partial J_3}{\partial w_1}\right] \tag{10}$$

in which the gradients of $J_t, t = 1,2,3$ with respect to w_1 is obtained as

$$\frac{\partial J_t}{\partial w_1} = \left[\bar{y}(t) - y(t)\right] \times \left.\frac{\partial y}{\partial w_1}\right|_{y=y(t)}, t = 1,2,3 \tag{11}$$

where

$$\frac{\partial y}{\partial w_1} = \frac{\partial z_2}{\partial w_1} = \frac{\partial z_2}{\partial f(z_1)} \frac{\partial f(z_1)}{\partial z_1} \frac{\partial z_1}{\partial w_1} \tag{12}$$

with

$$\frac{\partial z_2}{\partial f(z_1)} = \frac{\partial\left[w_2 f(z_1)\right]}{\partial f(z_1)} = w_2 \tag{13}$$

$$\frac{\partial f(z_1)}{\partial z_1} = f(z_1)(1 - f(z_1)) \tag{14}$$

$$\frac{\partial z_1}{\partial w_1} = \frac{\partial(w_1 x)}{\partial w_1} = x \tag{15}$$

Therefore,

$$\begin{cases} \frac{\partial J_1}{\partial w_1} = \left[\bar{y}(1) - y(1)\right]\bar{w}_2 f\left[\bar{z}_1(1)\right]\left[1 - f\left[\bar{z}_1(1)\right]\right]x(1) \\ = (0.0574 - 0.3) \times 0.1 \times 0.574 \times (1 - 0.574) \times 1 = -0.00593 \\ \frac{\partial J_1}{\partial w_2} = \left[\bar{y}(2) - y(2)\right]\bar{w}_2 f\left[\bar{z}_1(2)\right]\left[1 - f\left[\bar{z}_1(2)\right]\right]x(2) \\ = (0.0611 - 0.6) \times 0.1 \times 0.611 \times (1 - 0.611) \times 1.5 = -0.0192 \\ \frac{\partial J_1}{\partial w_3} = \left[\bar{y}(3) - y(3)\right]\bar{w}_2 f\left[\bar{z}_1(3)\right]\left[1 - f\left[\bar{z}_1(3)\right]\right]x(3) \\ = (0.0646 - 0.8) \times 0.1 \times 0.646 \times (1 - 0.646) \times 2 = -0.034 \end{cases} \tag{16}$$

Substituting (16) into gradient (10), yields

$$\frac{\partial J}{\partial w_1} = \frac{1}{3} \times (-0.00593 - 0.0192 - 0.034) = -0.0197 \tag{17}$$

(d):

The weights w_2 and w_1 can be updated by using the gradient decent approach as

$$\begin{aligned} w_2 &= \bar{w}_2 - \lambda \frac{\partial J}{\partial w_2} = 0.1 + 0.5 \times 0.314 = 0.257 \\ w_1 &= \bar{w}_1 - \lambda \frac{\partial J}{\partial w_1} = 0.3 + 0.5 \times 0.0197 = 0.31 \end{aligned} \tag{18}$$

~~Solution 2:~~

Step 1: Determine the weight matrices

$$\mathbf{W}^{[2]} = \begin{bmatrix} w_{0,1}^{[2]} & w_{0,2}^{[2]} & w_{0,3}^{[2]} \\ \vdots & \vdots & \vdots \\ w_{4,1}^{[2]} & w_{4,2}^{[2]} & w_{4,3}^{[2]} \end{bmatrix}, \mathbf{W}^{[1]} = \begin{bmatrix} w_{0,1}^{[1]} & \cdots & w_{0,4}^{[1]} \\ \vdots & \ddots & \vdots \\ w_{8,1}^{[1]} & \cdots & w_{8,4}^{[1]} \end{bmatrix}$$

Step 2: Calculate the output layer gradient

$$\nabla\mathbf{J}_{\mathbf{w}^{[2]}}=\begin{bmatrix}\frac{\partial J_1}{\partial w_{0,1}^{[2]}} & \cdots & \frac{\partial J_3}{\partial w_{0,4}^{[2]}}\\ \vdots & \ddots & \vdots\\ \frac{\partial J_1}{\partial w_{8,1}^{[2]}} & \cdots & \frac{\partial J_3}{\partial w_{8,4}^{[2]}}\end{bmatrix}=[(\hat{\mathbf{Y}}-\mathbf{Y})\mathbf{H}^{\mathrm{T}}]^{\mathrm{T}}$$

where

$$\hat{\mathbf{Y}}-\mathbf{Y}=\begin{bmatrix}\frac{\partial J_1}{\partial \hat{y}_1}\\ \frac{\partial J_2}{\partial \hat{y}_2}\\ \frac{\partial J_3}{\partial \hat{y}_3}\end{bmatrix},\ \mathbf{H}=\begin{bmatrix}\frac{\partial \hat{y}_k}{\partial w_{0,k}^{[2]}}\\ \vdots\\ \frac{\partial \hat{y}_k}{\partial w_{4,k}^{[2]}}\end{bmatrix}$$

Step 3: Calculate the input layer gradient

$$\nabla\mathbf{J}_{\mathbf{w}^{[1]}}=\begin{bmatrix}\sum_{k=1}^{3}\frac{\partial J_k}{\partial w_{0,1}^{[1]}} & \sum_{k=1}^{3}\frac{\partial J_k}{\partial w_{0,2}^{[1]}} & \sum_{k=1}^{3}\frac{\partial J_k}{\partial w_{0,3}^{[1]}}\\ \vdots & \vdots & \vdots\\ \sum_{k=1}^{3}\frac{\partial J_k}{\partial w_{4,1}^{[1]}} & \sum_{k=1}^{3}\frac{\partial J_k}{\partial w_{4,2}^{[1]}} & \sum_{k=1}^{3}\frac{\partial J_k}{\partial w_{4,3}^{[1]}}\end{bmatrix}$$

$$=\left(\{[(\hat{\mathbf{Y}}-\mathbf{Y})^{\mathrm{T}}(\bar{\mathbf{W}}^{[2]})^{\mathrm{T}}]^{\mathrm{T}}\odot\mathbf{H}_{\mathrm{d}}\}\mathbf{X}^{\mathrm{T}}\right)^{\mathrm{T}}$$

where

$$\hat{\mathbf{Y}}-\mathbf{Y}=\begin{bmatrix}\frac{\partial J_1}{\partial \hat{y}_1}\\ \frac{\partial J_2}{\partial \hat{y}_2}\\ \frac{\partial J_3}{\partial \hat{y}_3}\end{bmatrix},\ \bar{\mathbf{W}}^{[2]}=\begin{bmatrix}\frac{\partial \hat{y}_1}{\partial h_1} & \frac{\partial \hat{y}_2}{\partial h_1} & \frac{\partial \hat{y}_3}{\partial h_1}\\ \vdots & \vdots & \vdots\\ \frac{\partial \hat{y}_1}{\partial h_4} & \frac{\partial \hat{y}_2}{\partial h_4} & \frac{\partial \hat{y}_3}{\partial h_4}\end{bmatrix}=\begin{bmatrix}w_{0,1}^{[2]} & w_{0,2}^{[2]} & w_{0,3}^{[2]}\\ \vdots & \vdots & \vdots\\ w_{4,1}^{[2]} & w_{4,2}^{[2]} & w_{4,3}^{[2]}\end{bmatrix},$$

$$\mathbf{H}_{\mathrm{d}}=\begin{bmatrix}\frac{\partial h_1}{\partial z_1}\\ \vdots\\ \frac{\partial h_4}{\partial z_4}\end{bmatrix},\ \mathbf{X}=\begin{bmatrix}1\\ \vdots\\ x_8\end{bmatrix}$$