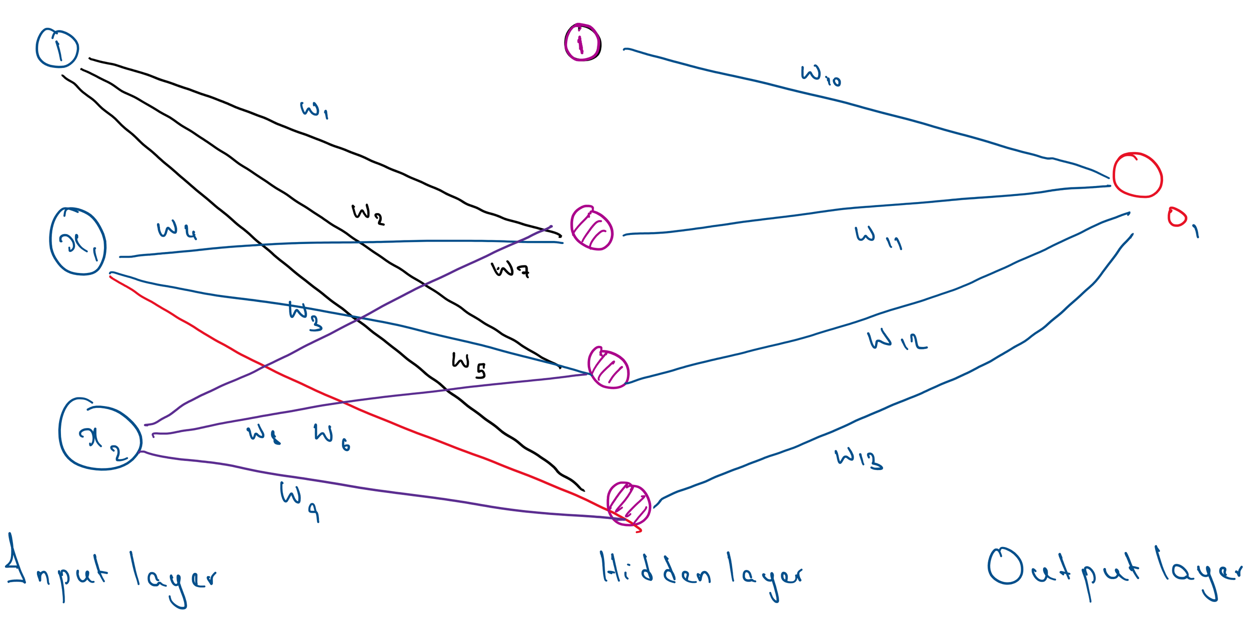
Data Analysis and Machine Learning: Problem Set 5

Tuesday, 29 March 2022 8:04 pm

Q1)



- Choose a sigmoid activation function

$$S(x) = \frac{1}{1+e^{-x}}$$

- Denote α to be the input of each neuron, H = hidden layer neurons and O = output layer neuron, β = output of each neuron.

$$\alpha_{H_1} = w_1 + x_1 w_4 + x_2 w_7 \qquad \beta_{H_1} = f(\alpha_{H_1})$$

$$\alpha_{H_2} = w_2 + x_1 w_5 + x_2 w_8 \qquad \beta_{H_2} = f(\alpha_{H_2})$$

$$\alpha_{H_3} = w_3 + x_1 w_6 + x_2 w_9 \qquad \beta_{H_3} = f(\alpha_{H_3})$$

- Output layer:

$$\underbrace{\beta_{O_1}}_{\text{predicted output value.}} = f(\alpha_{O_1}) = w_{10} + \beta_{H_1} w_{11} + \beta_{H_2} w_{12} + \beta_{H_3} w_{13}$$

- Define the initial loss function:

$$E = \frac{1}{2}\sum_i^N (y - y_{pred})^2 = \frac{1}{2}(y - \beta_{O_1})^2$$

- Suppose one wishes to compute the derivative of the loss function w.r.t. w_{10}:

$$\frac{\partial E}{\partial w_{10}} = \frac{\partial E}{\partial \beta_{O_1}} \frac{\partial \beta_{O_1}}{\partial \alpha_{O_1}} \frac{\partial \alpha_{O_1}}{\partial w_{10}}$$

$$\therefore \frac{\partial E}{\partial \beta_{O_1}} = -(y - \beta_{O_1}) \cdot \quad \frac{\partial \beta_{O_1}}{\partial \alpha_{O_1}} = \beta_{O_1}(1-\beta_{O_1}) \quad \frac{\partial \alpha_{O_1}}{\partial w_{10}} = 1$$

$$\therefore \frac{\partial E}{\partial w_{10}} = -\beta_{O_1}(y - \beta_{O_1})(1-\beta_{O_1})$$

- One can update the estimate of w_{10} as follows:

$$w_{10}^{(t+1)} = w_{10}^{(t)} - \underbrace{\eta}_{\text{learning rate}} \frac{\partial E}{\partial w_{10}}$$

$$= w_{10}^{(t)} + \eta \beta_{O_1}(y - \beta_{O_1})(1-\beta_{O_1})$$

$$\frac{\partial E}{\partial w_1} = \frac{\partial E}{\partial \beta_{O_1}} \frac{\partial \beta_{O_1}}{\partial \alpha_{O_1}} \cdot \frac{\partial \alpha_{O_1}}{\partial w_1}$$

$$= -(y - \beta_{O_1})(1-\beta_{O_1})\beta_{O_1} w_{11} \beta_{H_1}(1-\beta_{H_1})$$

$$\frac{\partial E}{\partial w_2} = \frac{\partial E}{\partial \beta_{O_1}} \frac{\partial \beta_{O_1}}{\partial \alpha_{O_1}} \frac{\partial \alpha_{O_1}}{\partial w_2}$$

$$= -(y - \beta_{O_1})(1-\beta_{O_1})\beta_{O_1} w_{12} \beta_{H_2}(1-\beta_{H_2})$$

$$\frac{\partial E}{\partial w_3} = \frac{\partial E}{\partial \beta_{O_1}} \frac{\partial \beta_{O_1}}{\partial \alpha_{O_1}} \frac{\partial \alpha_{O_1}}{\partial w_3}$$

$$= -(y - \beta_{O_1})(1-\beta_{O_1})\beta_{O_1} w_{13} \beta_{H_3}(1-\beta_{H_3})$$

$$\frac{\partial E}{\partial w_4} = \frac{\partial E}{\partial \beta_{O_1}} \frac{\partial \beta_{O_1}}{\partial \alpha_{O_1}} \frac{\partial \alpha_{O_1}}{\partial w_4}$$

$$-(y - \beta_{O_1})(1-\beta_{O_1})\beta_{O_1} w_{11} \beta_{H_1}(1-\beta_{H_1}) x_1$$

$$\frac{\partial E}{\partial w_5} = \frac{\partial E}{\partial \beta_{O_1}} \frac{\partial \beta_{O_1}}{\partial \alpha} \frac{\partial \alpha_{O_1}}{\partial w_5}$$

$$= -(y - \beta_{O_1})(1-\beta_{O_1})\beta_{O_1} \beta_{H_2} w_{12}(1-\beta_{H_2}) x_1$$

$$\frac{\partial E}{\partial w_6} = -(y - \beta_{O_1})(1-\beta_{O_1})\beta_{O_1} \beta_{H_3} w_{13}(1-\beta_{H_3}) x_1$$

$$\frac{\partial E}{\partial w_7} = -(y - \beta_{O_1})(1-\beta_{O_1})\beta_{O_1} \beta_{H_1} w_{11}(1-\beta_{H_1}) x_2$$

$$\frac{\partial E}{\partial w_8} = -(y - \beta_{O_1})(1-\beta_{O_1})\beta_{O_1} \beta_{H_2} w_{12}(1-\beta_{H_2}) x_2$$

$$\frac{\partial E}{\partial w_9} = -(y - \beta_{O_1})(1-\beta_{O_1})\beta_{O_1} \beta_{H_3} w_{13}(1-\beta_{H_3}) x_2$$

$$\frac{\partial E}{\partial w_{10}} \quad -(y - \beta_{O_1})(1-\beta_{O_1})$$

$$\frac{\partial E}{\partial w_{11}} = -(y - \beta_{O_1})(1-\beta_{O_1})\beta_{H_1}$$

$$\frac{\partial E}{\partial w_{12}} = -(y - \beta_{O_1})(1-\beta_{O_1})\beta_{H_2}$$

$$\frac{\partial E}{\partial w_{13}} = -(y - \beta_{O_1})(1-\beta_{O_1})\beta_{H_3}$$

- The general procedure for updating weights is given by:

$$w_n^{(t+1)} = w_n^{(t)} - \frac{\partial E}{\partial w_n}$$

$$\therefore w_1^{(t+1)} = w_1 + (y - \beta_{O_1})(1-\beta_{O_1})\beta_{O_1} w_{11}\beta_{H_1}(1-\beta_{H_1})$$

$$w_2^{(t+1)} = w_2 + (y - \beta_{O_1})(1-\beta_{O_1})\beta_{O_1} w_{12}\beta_{H_2}(1-\beta_{H_2})$$

$$w_3^{(t+1)} = w_3 + (y - \beta_{O_1})(1-\beta_{O_1})\beta_{O_1} w_{13}\beta_{H_3}(1-\beta_{H_3})$$

$$w_4^{(t+1)} = w_4 + (y - \beta_{O_1})(1-\beta_{O_1})\beta_{O_1} w_{11}\beta_{H_1}(1-\beta_{H_1}) x_1$$

$$w_5^{(t+1)} = w_5 + (y - \beta_{O_1})(1-\beta_{O_1})\beta_{O_1} \beta_{H_2} w_{12}(1-\beta_{H_2}) x_1$$

$$w_6^{(t+1)} = w_6 + (y - \beta_{O_1})(1-\beta_{O_1})\beta_{O_1} \beta_{H_3} w_{13}(1-\beta_{H_3}) x_1$$

$$w_7^{(t+1)} = w_7 + (y - \beta_{O_1})(1-\beta_{O_1})\beta_{O_1} \beta_{H_1} w_{11}(1-\beta_{H_1}) x_2$$

$$w_8^{(t+1)} = w_8 + (y - \beta_{O_1})(1-\beta_{O_1})\beta_{O_1} \beta_{H_2} w_{12}(1-\beta_{H_2}) x_2$$

$$w_9^{(t+1)} = w_9 + (y - \beta_{O_1})(1-\beta_{O_1})\beta_{O_1} \beta_{H_3} w_{13}(1-\beta_{H_3}) x_2$$

$$w_{10}^{(t+1)} = w_{10} + (y - \beta_{O_1})(1-\beta_{O_1})$$

$$w_{11}^{(t+1)} = w_{11} + (y - \beta_{O_1})(1-\beta_{O_1})\beta_{H_1}$$

$$w_{12}^{(t+1)} = w_{12} + (y - \beta_{O_1})(1-\beta_{O_1})\beta_{H_1}$$

$$w_{13}^{(t+1)} = w_{13} + (y - \beta_{O_1})(1-\beta_{O_1})\beta_{H_3}$$