Assignment 2

Tuesday, March 12, 2024 7:39 PM

2 layers

$$L = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{Y}_i - Y_i\right)^2$$

Layer 1 input: $z_1 = XW_1 + b_1$

Layer 1 activation: $A_1 = \sigma(z_1) = \frac{1}{1+e^{-z_1}}$

Layer 2 input: $z_2 = A_1W_2 + b_2$

Output: $\hat{y} = z_2$ ← linear function for output

$\hat{y}$ → y_pred

$\sigma'(x) = \sigma(x)(1-\sigma(x))$

Backprop

Gradient of output wrt. z_2

$$\frac{\partial L}{\partial z_2} = \frac{2}{n}(\hat{y} - y)$$

Gradient wrt. W_2

$$\frac{\partial L}{\partial W_2} = A_1^T \cdot \frac{\partial L}{z_2}$$

Gradient wrt b_2

$$\frac{\partial L}{\partial b_2} = 1$$

Gradient of z_2 wrt. W_2

$$\frac{\partial z_2}{\partial W_2} = A_1 \quad b/c \quad z_2 = A_1W_2 + b_2$$

Back prop to hidden layer

$$\frac{\partial L}{\partial A_1} = \frac{\partial L}{\partial z_2} \cdot W_2^T \qquad \frac{\partial L}{\partial z_1} = \frac{\partial L}{\partial A_1} \cdot \sigma'(z_1)$$

$$\frac{\partial L}{\partial W_1} = X^T \cdot \frac{\partial L}{\partial z_1}$$

Output

$$W_2 = W_2 - \alpha \cdot \partial L / \partial W_2$$

$$b_2 = b_2 - \alpha \cdot \partial L / \partial z_2$$

Hidden

$$W_1 = W_1 \cdot \alpha\, \partial L / \partial W_1$$

$$b_1 = b_1 \cdot \alpha\, \partial L / \partial b_1$$