An Overview on Deep Neural Networks: Part 3

Activation Functions' Derivatives

Activation Functions' Derivatives

$$\sigma'^{[l]}\left(Z^{[l]}\right)_{i,j,u,v} = \frac{\partial}{\partial Z_{u,v}^{[l]}} \sigma^{[l]}\left(Z^{[l]}\right)_{i,j}$$

ReLU

$$\sigma^{[l]}\left(Z^{[l]}\right)_{i,j} = \max\left\{0, Z_{i,j}^{[l]}\right\}$$

$$\sigma'^{[l]}\left(Z^{[l]}\right)_{i,j,u,v} = \begin{cases} 1 & i = u \text{ and } j = v \text{ and } Z_{u,v}^{[l]} > 0 \\ 0 & \text{otherwise} \end{cases}$$

Sigmoid

$$\sigma^{[l]}\left(Z^{[l]}\right)_{i,j} = \frac{1}{1 + \exp\left(-Z_{i,j}^{[l]}\right)}$$

$$\sigma'^{[l]}\left(Z^{[l]}\right)_{i,j,u,v} = \begin{cases} A_{i,j}^{[l]}\left(1 - A_{i,j}^{[l]}\right) & i = u \text{ and } j = v \\ 0 & \text{otherwise} \end{cases}$$

Softmax

$$\sigma^{[l]}\left(Z^{[l]}\right)_{i,j} = \frac{\exp\left(Z_{i,j}^{[l]}\right)}{\sum_{k=1}^{m_h^{[l]}} \exp\left(Z_{i,k}^{[l]}\right)}$$

$$\sigma'^{[l]}\left(Z^{[l]}\right)_{i,j,u,v} = \begin{cases} A_{i,j}^{[l]}\left(1 - A_{i,j}^{[l]}\right) & i = u \text{ and } j = v \\ -A_{i,j}^{[l]} A_{i,v}^{[l]} & i = u \text{ and } j \neq v \\ 0 & i \neq u \end{cases}$$

Loss Functions’ Derivatives

SSE

$$J(Y,\hat{Y}) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m_y}\left(\hat{Y}_{i,j} - Y_{i,j}\right)^2$$

$$\frac{\partial J}{\partial \hat{Y}_{i,j}} = \hat{Y}_{i,j} - Y_{i,j}$$

$$\frac{\partial J}{\partial \hat{Y}} = \hat{Y} - Y$$

BCE

$$J\left(Y,\hat{Y}\right) = -\sum_{i=1}^{n}\sum_{j=1}^{m_y}\left(Y_{i,j}\log\left(\hat{Y}_{i,j}\right) + \left(1 - Y_{i,j}\right)\log\left(1 - \hat{Y}_{i,j}\right)\right)$$

$$\frac{\partial J}{\partial \hat{Y}_{i,j}} = \frac{\hat{Y}_{i,j} - Y_{i,j}}{\hat{Y}_{i,j}\left(1 - \hat{Y}_{i,j}\right)}$$

$$\frac{\partial J}{\partial \hat{Y}} = \left(\hat{Y} - Y\right) \oslash \left(\hat{Y} \odot \left(1 - \hat{Y}\right)\right)$$

CCE

$$J(Y, \hat{Y}) = -\sum_{i=1}^{n}\sum_{j=1}^{m_y}\left(Y_{i,j}\log\left(\hat{Y}_{i,j}\right)\right)$$

$$\frac{\partial J}{\partial \hat{Y}_{i,j}} = -\frac{Y_{i,j}}{\hat{Y}_{i,j}}$$

$$\frac{\partial J}{\partial \hat{Y}} = -Y \oslash \hat{Y}$$

Different Types of Gradient Descent

Different Types of Gradient Descent

- Batch
- Stochastic
- Mini-Batch

Some Mathematical Notations and Their NumPy Equivalents

Some Mathematical Notations and Their NumPy Equivalents

AB	$\equiv$	`A @ B`
$A : B$	$\equiv$	`np.tensordot(A, B)`
$A \odot B$	$\equiv$	`A * B`
$A \oslash B$	$\equiv$	`A / B`