Introduction to Statistical Learning

Omid Safarzadeh

January 19, 2022

Table of Contents

Exponential Linear Unit (ELU)

$$ELU(x) = \begin{cases} x & x \geq 0 \\ \\ \alpha(e^x - 1) & x < 0 \end{cases}$$

$$ELU'(x) = \begin{cases} 1 & x < 0 \\ \\ \alpha e^x & x < 0 \end{cases}$$

Exponential activation function

$$exp(x) = exp'(x) = x$$

Gaussian error linear unit (GELU)

$$GELU(x) = xP(X \leq x) = x\Phi(x)$$

$$\approx 0.5x(1 + \tanh\left[\sqrt{2/\pi}(x + 0.044715x^3)\right]$$

$$GELU'(x) \approx 0.5(\tanh\left[\sqrt{2/\pi}(0.044715x^3 + x)\right] + 1)$$

$$+\frac{1}{\sqrt{2\pi}}[x(0.134145x^2 + 1)\, sech^2[\sqrt{2/\pi}(0.044715x^3)]$$

Hard sigmoid

$$f(x) = \min(1, \max(0, \frac{(2x+5)}{10}))$$

$$f'(x) = \begin{cases} 0 & \text{if } |x| > 2.5 \\ \\ 0.2 & \text{otherwise} \end{cases}$$

Rectified Linear Unit (ReLU)

$$ReLU(x) = \max(0, x)$$

$$ReLU'(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Scaled Exponential Linear Unit (SELU)

$$SELU(x) = \begin{cases} x & x > 0 \\ \\ \alpha(e^x - 1) & x \leq 0 \end{cases}$$

$$SELU'(x) = \begin{cases} 1 & x > 0 \\ \\ \alpha e^x & x < 0 \end{cases}$$

Sigmoid

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

$$Sigmoid'(x) = \frac{e^{x}}{(1 + e^{x})^{2}}$$

Softplus

$$Softplus(x) = \log(e^x + 1)$$

$$Softplus'(x) = \frac{e^x}{1 + e^x}$$

Softsign

$$Softsign(x) = \frac{x}{|x| + 1}$$

$$Softsign'(x) = \frac{1}{(|x| + 1)^2}$$

Swish

$$Swish(x) = x * Sigmoid(x) = \frac{x}{1 + e^{-x}}$$

$$Swish'(x) = \frac{1}{1 + e^{-x}} + \frac{xe^{-x}}{(1 + e^{-x})^2}$$

Hyperbolic Tangent

$$\tanh(x) = \frac{\sinh(x)}{\cosh x} = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

$$\tanh'(x) = \left(\frac{2e^x}{1 + e^{2x}}\right)^2$$

Mean Absolute Error (MAE)

$$MAE = \frac{1}{n}\Sigma_{i=0}^{n}|y_{-}\hat{y}_i|$$

Where $\hat{y}$ is the predicted value

Mean Absolute Percentage Error (MAPE)

$$MAPE = \frac{100\%}{n} \sum_{i=0}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

Where $\hat{y}$ is the predicted value

Mean Squared Error (MSE)

$$MSE = \frac{1}{n}\Sigma_{i=0}^{n}(y - \hat{y}_i)^2$$

Where $\hat{y}$ is the predicted value

Mean Squared Logarithmic Error (MSLE)

$$MSE = \frac{1}{n}\Sigma_{i=0}^{n}(\log(y_i + 1) - \log(\hat{y}_i + 1))^2$$

Where $\hat{y}$ is the predicted value

Indicator function

$$I(X \in A) := \begin{cases} 1 & \text{if} \quad X \in A \\ 0 & \text{otherwise} \end{cases}$$

- $E[I(X \in A)] = P(X \in A)$