# Introduction to Statistical Learning

Omid Safarzadeh

February 2, 2022

#### Table of Contents

- Activation Functions
  - Exponential Linear Unit (ELU)
  - Exponential activation function
  - Gaussian error linear unit (GELU)
  - Hard sigmoid
  - Rectified Linear Unit (ReLU)
  - Scaled Exponential Linear Unit (SELU)
  - Sigmoid
  - Softplus
  - Softsign
  - Swish
  - Hyperbolic Tangent
- 2 Loss Functions
  - Mean Absolute Error (MAE)
  - Mean Absolute Percentage Error (MAPE)
  - Mean Squared Error (MSE)
  - Mean Squared Error (MSE)
  - Indicator function



Omid Safarzadeh

# Exponential Linear Unit (ELU)

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

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

Omid Safarzadeh Introduction to Constitution Learning February 2, 2022

### Exponential activation function

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

Omid Safarzadeh Introduction to Statistical Learnin

## Gaussian error linear unit (GELU)

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

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

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

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

Omid Safarzadeh Installiason to Statistical Resolution February 2, 2022 5/18

### Hard sigmoid

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

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

Omid Safarzadeh Introduction to Statistical Learning February 2, 2022

## Rectified Linear Unit (ReLU)

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

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

Omid Safarzadeh Businesia Sanasia Lagara February 2, 2022 7/18

# Scaled Exponential Linear Unit (SELU)

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

Omid Safarzadeh Issochustion to Stanistical Issaming February 2, 2022 8

# Sigmoid

$$Sigmoid(x) = rac{1}{1 + e^{-x}}$$
  $Sigmoid'(x) = rac{e^{x}}{(1 + e^{x})^{2}}$ 

### Softplus

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



Omid Safarzadeh Introduction so Statistical Learning February 2, 2022

# Softsign

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



Omid Safarzadeh Ingreduction to September 1997 February 2, 2022

#### Swish

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



Omid Safarzadeh Introduction to Statistical Learning February 2, 2022

# Hyperbolic Tangent

$$\tanh(x) = \frac{\sinh(x)}{\cosh x} = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$$
$$\tanh'(x) = (\frac{2e^{x}}{1 + e^{2x}})^{2}$$

Omid Safarzadeh III/rosussous oscusionel Isosanius February 2, 2022

# Mean Absolute Error (MAE)

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

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



Omid Safarzadeh Introduction to Statistical Learning. February 2, 2022

## Mean Absolute Percentage Error (MAPE)

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

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

( D ) ( B ) ( E ) ( E ) ( C )

Omid Safarzadeh Introduction to Steutenest Learning February 2, 2022

# Mean Squared Error (MSE)

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

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



Omid Safarzadeh Untroduction to Statistical Learning February 2, 2022

## Mean Squared Logarithmic Error (MSLE)

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

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



Omid Safarzadeh Incoduction to Statistical Learning February 2, 2022

#### Indicator function

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

•  $E[I(X \in A)] = P(X \in A)$ 



Omid Safarzadeh Involución vo Statistical Iscamina February 2, 2022