Lecture 2.1

- Classification, Logistic Regression

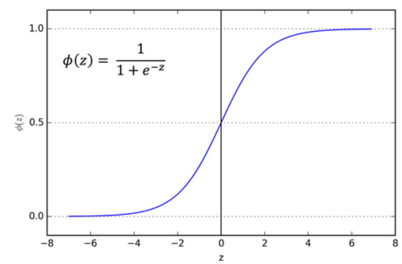
Classification and Logistic Regression

- **Classification** is a task or problem in machine learning where the goal is to assign input data to one of several predefined categories (classes)
- **Logistic Regression** (Binary Classification) is a statistical model used to predict the probability of a binary outcome (two possible classes, such as 0 and 1, Yes and No, True and False)

Key Concepts of Logistic Regression

- **Binary Output**: **Predicts** probabilities between **0** and **1** for two classes
- **Logistic Function (Sigmoid)**: Used to map predictions to probabilities

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$



where $z = w_0 + w_1x_1 + \quad + \cdots + w_nx_n$

- **Decision Boundary**: A threshold (e.g., 0.5) is used to classify probabilities into labels:

$$y = \begin{cases} 1 & \sigma(z) \geq 0.5 \\ 0 & \sigma(z) < 0.5 \end{cases}$$

Logistic Regression Equation

- The logistic regression equation is:

$$P(y = 1|x) = \frac{1}{1 + e^{-(w_0 + w_1 x)}}$$

Where,

$P(y = 1 \mid x)$: Probability that y=1 based on x

w_0: y-intercept

w_1: coefficient for x

Derivative of Sigmoid Function

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

$$\frac{d\big(\sigma(\mathrm{x})\big)}{dx} = \frac{0 * (1+e^{-\mathrm{x}}) - (1) * \big(e^{-\mathrm{x}} * (-1)\big)}{(1+e^{-\mathrm{x}})^2}$$

$$\frac{d\big(\sigma(\mathrm{x})\big)}{dx} = \frac{(e^{-\mathrm{x}})}{(1+e^{-\mathrm{x}})^2} = \frac{1\text{-}1 + (e^{-\mathrm{x}})}{(1+e^{-\mathrm{x}})^2} = \frac{1+e^{-\mathrm{x}}}{(1+e^{-\mathrm{x}})^2} - \frac{1}{(1+e^{-\mathrm{x}})^2}$$

$$\frac{d\big(\sigma(\mathrm{x})\big)}{dx} = \frac{1}{1+e^{-\mathrm{x}}} * \left(1 - \frac{1}{1+e^{-\mathrm{x}}}\right) = \sigma(\mathrm{x})\big(1-\sigma(\mathrm{x})\big)$$

Loss orCost Function

- The loss function is the product of likelihood for all observations:

$$L = \prod_{i=1}^{N} \sigma(z_i)^{y_i}[1 - \sigma(z_i)]^{1-y_i}$$

- Log-likelihood:

$$\log L = \sum_{i=1}^{N} [y_i \log(\sigma(z_i)) + (1 - y_i)\log(1 - \sigma(z_i))]$$

- Loss or CostFunction:

$$E(w) = -\frac{1}{N}\sum_{i=1}^{N} [y_i \log(\sigma(z_i)) + (1 - y_i)\log(1 - \sigma(z_i))]$$

Derivative of single term 1

$$e_i = -\left[y_i \log(\sigma(z_i)) + (1 - y_i) \log(1 - \sigma(z_i))\right]$$

Take the derivative with respect to w_0. Start by expanding $\sigma(z_i)$:

$$\sigma(z_i) = \frac{1}{1 + e^{-z_i}},$$

where $z_i = w_0 + w_1 x_i$.

Apply the Chain Rule

$$\frac{\partial e_i}{\partial w_0} = -\frac{\partial}{\partial w_0}\left[y_i \underbrace{\log(\sigma(z_i))}_{\text{First Term}} + \underbrace{(1 - y_i) \log(1 - \sigma(z_i))}_{\text{Second Term}}\right]$$

Derivative of single term 2

- First Term:

$$\frac{\partial}{\partial w_0}\left[y_i \log(\sigma(z_i))\right] = y_i \cdot \frac{1}{\sigma(z_i)} \cdot \frac{\partial \sigma(z_i)}{\partial w_0}.$$

- Second Term:

$$\frac{\partial}{\partial w_0}\left[(1-y_i)\log(1-\sigma(z_i))\right] = (1-y_i) \cdot \frac{1}{1-\sigma(z_i)} \cdot \frac{\partial(-\sigma(z_i))}{\partial w_0}.$$

- As already known: $\dfrac{\partial \sigma(z_i)}{\partial w_0} = \sigma(z_i)(1-\sigma(z_i))$

Derivative of single term 3

$$\frac{\partial e_i}{\partial w_0} = -\left[y_i \cdot \frac{\sigma(z_i)(1-\sigma(z_i))}{\sigma(z_i)} + (1-y_i) \cdot \frac{-\sigma(z_i)(1-\sigma(z_i))}{1-\sigma(z_i)}\right]$$

$$\frac{\partial e_i}{\partial w_0} = \sigma(z_i) - y_i.$$

Sum Over All Data Points

$$\frac{\partial E}{\partial w_0} = \frac{1}{N}\sum_{i=1}^{N}(\sigma(z_i) - y_i).$$

Gradient Descent with respect to w_1

- As known:

$$\frac{\partial \sigma(z_i)}{\partial z_i} = \sigma(z_i)(1 - \sigma(z_i)).$$

- And $z_i = w_0 + w_1 x_i,$ $\quad \frac{\partial z_i}{\partial w_1} = x_i.$

- Therefore, $\frac{\partial \sigma(z_i)}{\partial w_1} = \sigma(z_i)(1 - \sigma(z_i))x_i.$

$$\frac{\partial E}{\partial w_1} = \frac{1}{N}\sum_{i=1}^{N}(\sigma(z_i) - y_i)x_i$$

Gradient Descent with respect to w_0

$$\frac{\partial E}{\partial w_0} = \frac{1}{N} \sum_{i=1}^{N} (\sigma(z_i) - y_i).$$

$$w_j = w_j - \alpha \cdot \frac{\partial E}{\partial w_j}$$

Example 1

x	y
2	0
3	0
5	1

$$w_0 = 0, w_1 = 0, \alpha = 0.1$$

$$\frac{\partial E}{\partial w_1} = \frac{1}{N} \sum_{i=1}^{N} (\sigma(z_i) - y_i) x_i$$

$$\frac{\partial E}{\partial w_0} = \frac{1}{N} \sum_{i=1}^{N} (\sigma(z_i) - y_i).$$

Iteration 1

$$z = w_0 + w_1 x_i = 0 + 0.x_i = 0$$

$$\frac{\partial E}{\partial w_0} = \frac{1}{N}\sum_{i=1}^{N}(\sigma(z_i) - y_i) \quad = \frac{1}{3}[0.5 + 0.5 - 0.5] = 0.167$$

$$\frac{\partial E}{\partial w_1} = \frac{1}{N}\sum_{i=1}^{N}(\sigma(z_i) - y_i)x_i \quad = \frac{1}{3}[0.5 \times 2 + 0.5 \times 3 - 0.5 \times 5] = 0.0$$

$$w_j = w_j - \alpha \cdot \frac{\partial E}{\partial w_j}$$

<table>
<tr><th>x</th><th>y</th><th>z</th><th>$\sigma(z)$</th><th>$\sigma(z) - y$</th><th>w_0</th><th>w_1</th></tr>
<tr><td>2</td><td>0</td><td>0</td><td>0.5</td><td>0.5</td><td rowspan="3">0-
0.1*0.167
=-0.0167</td><td rowspan="3">0.00-
0.1*0.0=
0.0</td></tr>
<tr><td>3</td><td>0</td><td>0</td><td>0.5</td><td>0.5</td></tr>
<tr><td>5</td><td>1</td><td>0</td><td>0.5</td><td>-0.5</td></tr>
</table>

Iteration 2

$$\frac{\partial E}{\partial w_0} = \frac{1}{N}\sum_{i=1}^{N}(\sigma(z_i) - y_i) \quad = \frac{1}{3}[0.496 + 0.496 - 0.504] = 0.162$$

$$\frac{\partial E}{\partial w_1} = \frac{1}{N}\sum_{i=1}^{N}(\sigma(z_i) - y_i)x_i \quad = \frac{1}{3}[0.496 \times 2 + 0.496 \times 3 - 0.504 \times 5] = -0.04$$

x	y	z	$\sigma(z)$	$\sigma(z) - y$	w_0	w_1
2	0	0.0167	0.496	0.496	0.0167-0.1*0.162 =0.0005	0.00-0.1*-0.04 =0.004
3	0	0.0167	0.496	0.496		
5	1	0.0167	0.496	-0.504		