TAMS ARTIFICIAL INTELLIGENCE SOCIETY
2021 - 2022

Logistic Regression

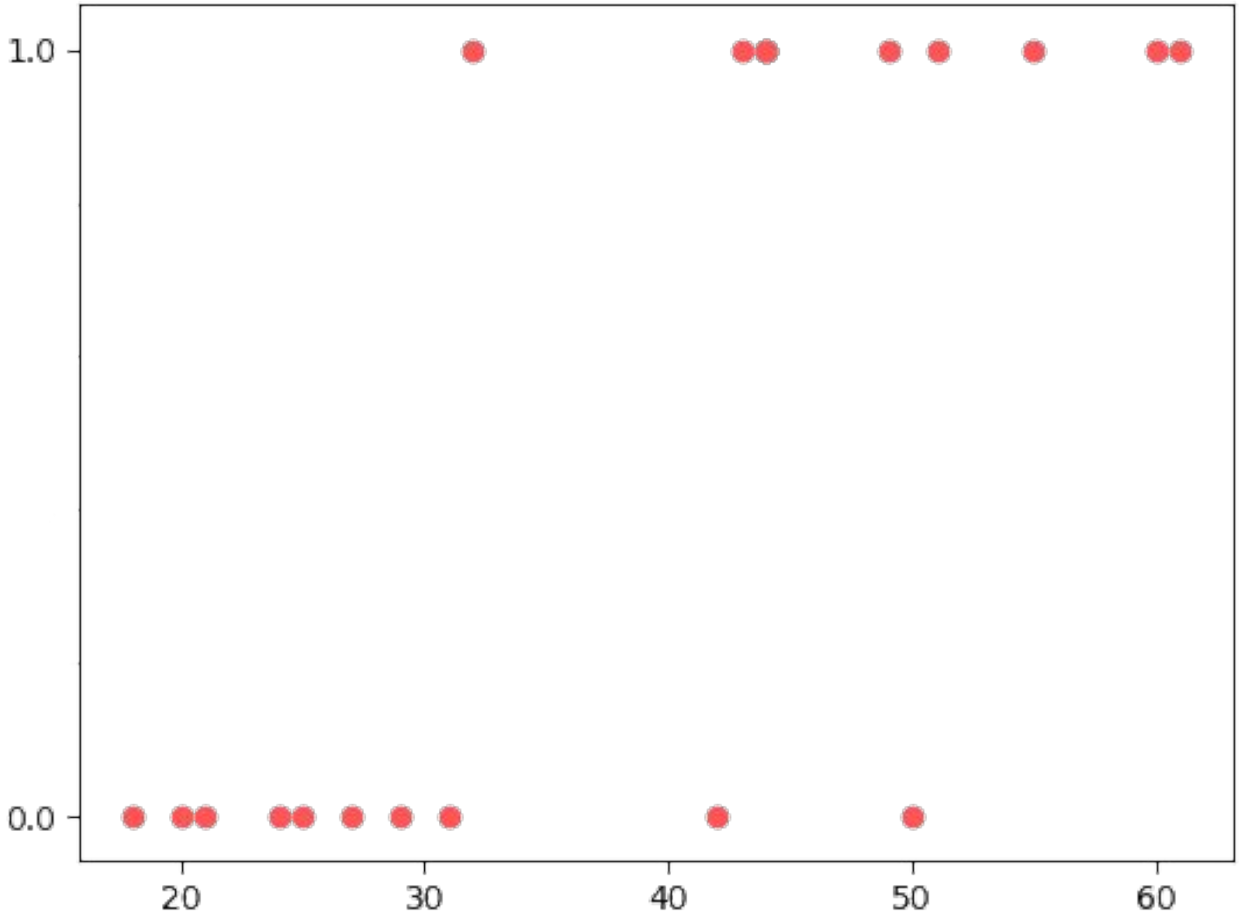
9/29/2021

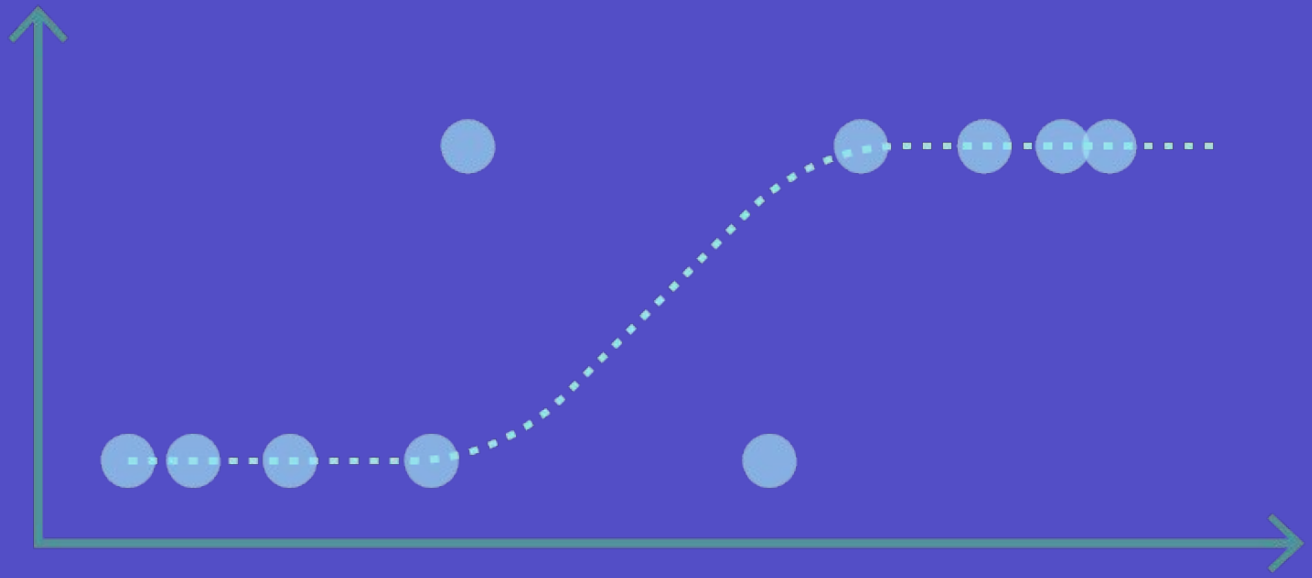
Lesson Plan

- Purpose
- Sigmoid Function
- Cost Function
- Gradient Descent

Purpose

- Used mainly for classification
- Example: is an email spam or not (binary)
- Finds $P(y = 1 \mid x ; \theta)$
- Output: either 0 or 1
- Can also be multiclass classification


1.0
0.0
20
30
40
50
60



Sigmoid Function

$$h_\theta(\mathbf{x}) = g(\theta^\top \mathbf{x})$$

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

Cost Function

$$Cost(h_\theta(x), y) = \begin{cases} -log(h_\theta(x)) & \text{if } y = 1 \\ -log(1 - h_\theta(x)) & \text{if } y = 0 \end{cases}$$

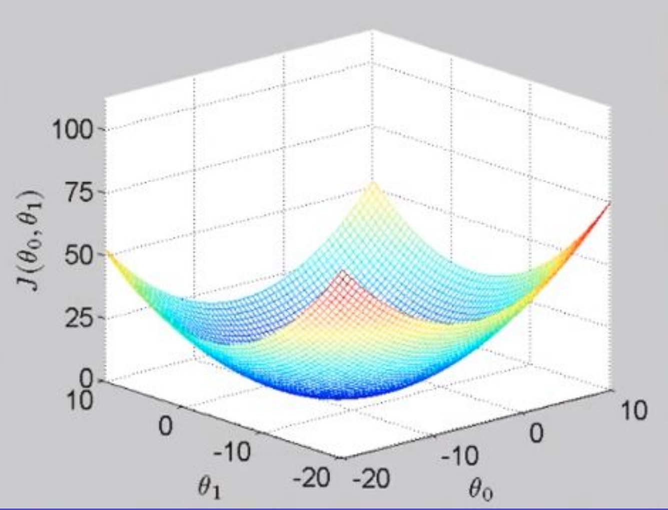
$$Cost(h_\theta(x), y) = -ylog(h_\theta(x)) - (1 - y)log(h_\theta(x))$$

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m} Cost(h_\theta(x^{(i)}), y^{(i)})$$

$$J(\theta) = \frac{1}{m}[\sum_{i=1}^{m} -y^{(i)}log(h_\theta(x^{(i)})) + (1 - y^{(i)})log(1 - h_\theta(x^{(i)}))]$$

$$m = number\ of\ samples$$

Gradient Descent



$$\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

- ***α***: learning rate
- Low ***α*** → slow convergence
- High ***α*** → may cause divergence

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

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

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

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

$$J(\theta) = \frac{1}{m}[\sum_{i=1}^{m} -y^{(i)} log(h_\theta(x^{(i)})) + (1 - y^{(i)}) log(1 - h_\theta(x^{(i)}))]$$

$$\frac{\partial(J(\theta))}{\partial(\theta j)} = -\frac{1}{m} * \sum_{i=1}^{m} \left[y^{(i)} * \frac{1}{h_\theta(x^{(i)})} * \frac{\partial\left(h_\theta(x^{(i)})\right)}{\partial(\theta j)} \right] + \sum_{i=1}^{m} \left[\left(1 - y^{(i)}\right) * \frac{1}{\left(1 - h_\theta(x^{(i)})\right)} * \frac{\partial\left(1 - h_\theta(x^{(i)})\right)}{\partial(\theta j)} \right]$$

$$\frac{\partial(J(\theta))}{\partial(\theta j)} = -\frac{1}{m} * (\sum_{i=1}^{m} \left[y^{(i)} * \frac{1}{h_\theta(x^{(i)})} * \sigma(z)\left(1 - \sigma(z)\right) * \frac{\partial(\theta^T x)}{\partial(\theta j)} \right] + \sum_{i=1}^{m} \left[\left(1 - y^{(i)}\right) * \frac{1}{\left(1 - h_\theta(x^{(i)})\right)} * (-\sigma(z)\left(1 - \sigma(z)\right) * \frac{\partial(\theta^T x)}{\partial(\theta j)} \right])$$

$$\frac{\partial(J(\theta))}{\partial(\theta j)} = -\frac{1}{m}*(\sum_{i=1}^{m}\left[y^{(i)}\frac{1}{h_\theta(x^{(i)})}h_\theta(x^{(i)})\left(1-h_\theta(x^{(i)})\right)*x_j^i\right]+$$

$$\sum_{i=1}^{m}\left[\left(1-y^{(i)}\right)*\frac{1}{\left(1-h_\theta(x^{(i)})\right)}*\left(-h_\theta(x^{(i)})\left(1-h_\theta(x^{(i)})\right)*x_j^i\right])$$

$$\frac{\partial(J(\theta))}{\partial(\theta j)} = -\frac{1}{m}*(\sum_{i=1}^{m}\left[y^{(i)}*\left(1-h_\theta(x^{(i)})\right)*x_j^i - \left(1-y^{(i)}\right)*h_\theta(x^{(i)})**x_j^i\right]$$

$$\frac{\partial(J(\theta))}{\partial(\theta j)} = -\frac{1}{m}*(\sum_{i=1}^{m}[\, y^{(i)}-y^{(i)}*h_\theta(x^{(i)})-h_\theta(x^{(i)})+y^{(i)}*h_\theta(x^{(i)})]*x_j^i\,)$$

$$\frac{\partial(J(\theta))}{\partial(\theta j)} = -\frac{1}{m}*(\sum_{i=1}^{m}[\, y^{(i)}-h_\theta(x^{(i)})]*x_j^i\,)$$

Parameter Update

$$\theta_j^{\text{new}} = \theta_j^{\text{old}} - \alpha \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) x_j^{(i)} \text{ for } j = 0, 1, \ldots, n$$