Machine Learning Foundations

Logistic Regression

Mingshan
August 20, 2020

Table of Contents

Two types of prediction

- The predict target is **continuous** valued → **Regression** problem

  $y \in [5, 50]$ the boston house-prices dataset

- The predict target is **discrete** valued → **Classification** problem

  $y \in \{0, 1\}$ the breast cancer wisconsin dataset

  $y \in \{0, 1, 2\}$ the iris dataset

Sigmoid Function

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

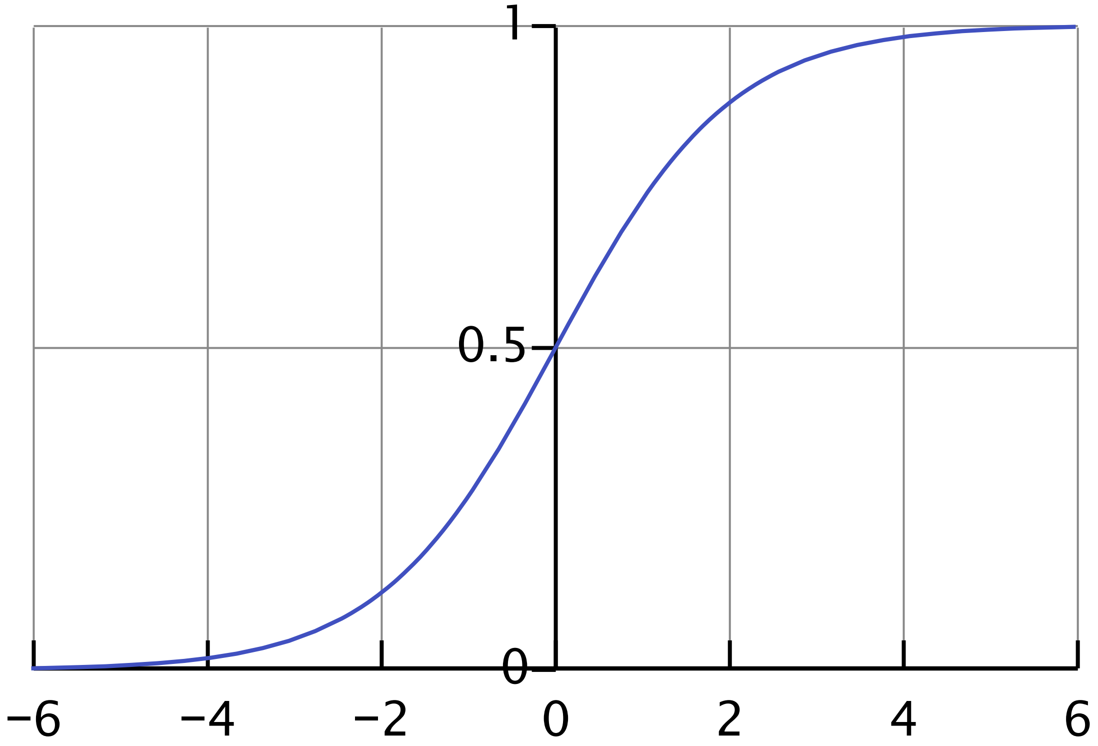
previously,we have:

$h_\theta(x) = \theta^T x$

Now,

$h_\theta(x) = \sigma(\theta^T x)$

$$h_\theta(x) = \frac{1}{1+\ e^{-\theta^T x}}$$



Meaning of the hypothesis in Logistic Regression

$h_\theta(x) = P(y = 1 | x; \theta)$ probability that y = 1, given x, parameterized by θ

$P(y = 0 | x; \theta) = 1 - P(y = 1 | x; \theta)$

Decision Boundary

- $h_\theta(x) = 0.5$ ➔ $\theta^T x = 0$

- It is a line that separates y = 1 area and y = 0 area.
- It is determined by the parameters

Logistic regression model

A modified hypothesis

+

A decision boundary

$$h_\theta(x) = \theta^T x$$

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}}$$

$$h_\theta(x) \geq 0.5$$

$$h_\theta(x) \geq 0.5 \Rightarrow y = 1$$
$$h_\theta(x) < 0.5 \Rightarrow y = 0$$

or,

$$\theta^T x \geq 0$$

$$\theta^T x \geq 0 \Rightarrow y = 1$$
$$\theta^T x < 0 \Rightarrow y = 0$$

Cost Function

- In linear regression model, we have defined a cost function:

$$J(\theta) = \frac{1}{2m}\sum_{i=1}^{m}(h_\theta\left(x^{(i)}\right) - y^{(i)})^2$$

- Can we continue to this?

Cost Function of logistic regression

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

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

Gradient Descent for Logistic Regression

- $Minimize\ J(\theta)\ over\ \theta$

- repeat until $J(\theta)$ converges to minimum:

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

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

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

Use the learned model to predict

- $h_\theta(x) = \frac{1}{1+ e^{-\theta^T x}}$

- $\hat{y} = \begin{cases} 1, & if\ h_\theta(x) \geq 0.5 \\ 0, & if\ h_\theta(x) < 0.5 \end{cases}$