Advanced Machine learning Mastering Course

Introduced by

George Samuel

Master in computer science
Cairo University

Innovisionray.com

2024

Agenda

Three Types of machine learning models

Ml Models

Geometric models	Linear Regression
Logical models	Decision Tree
Probabilistic Models	Logistic Regression

CLASSIFICATION

The classification problem is just like the regression problem, except that the values **y** we now want to predict take on only a **small number of discrete values**.

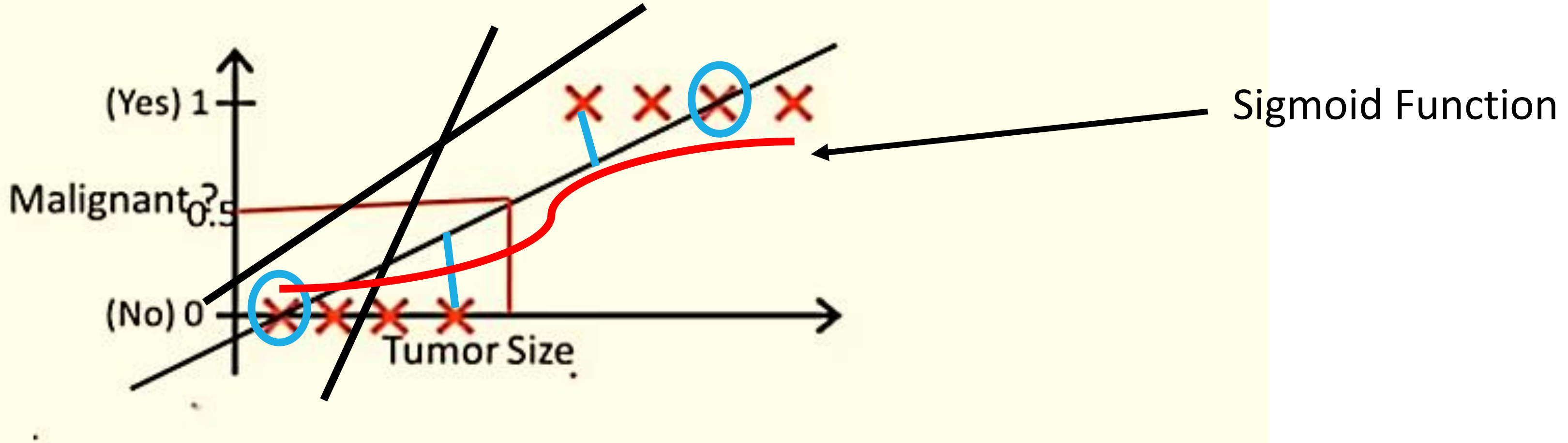
Some Example of Classification problem

- Email : Spam / Not spam
- Tumor: Malignant/ Benign

$y \in \{0, 1\}$

0: "Negative Class" (e.g., benign tumor)
1: "Positive Class" (e.g., malignant tumor)

Logistic Regression



Threshold classifier output $h_\theta(x)$ at 0.5:

If $h_\theta(x) \geq 0.5$, predict "y = 1"

If $h_\theta(x) < 0.5$, predict "y = 0"

Logistic Regression

- Binary Logistic Regression
  - We have a set of feature vectors X with corresponding binary outputs

$$X = \{x1, x2, \ldots, x_n\}^T$$

$$Y = \{y1, y2, \ldots, y_n\}^T, where \quad y_i \in \{0,1\}$$

  - We want to model p(y|x)

$$p(y_i = 1 \mid x_i, \theta) = \sum_j \theta_j x_{ij} \quad = x_i\theta$$

By definition $p(y_i = 1 \mid x_i, \theta) \in \{0,1\}$. We want to transform the probability to remove the range restrictions, as $x_i\theta$ can take any real value.

Logistic Regression

- **Hypothesis** $h_\theta(x) = g(\theta^T x)$

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

- **Interpretation**

$$h_\theta(x) = p(y = 1|x, \theta)$$

- **Because probabilites should sum to 1, define**

$$p(y = 0|x, \theta) := 1 - p(y = 1|x, \theta)$$

- **If $h_\theta(x) = 0.7$ interpret as 70% chance data point belongs to class**
- **If $h_\theta(x) \geq 0.5$ classify as *positive sentiment, malignant tumor, …***

Logistic Regression

- Odds

  p : probability of an event occurring

  1 – p : probability of the event not occurring

  The odds for event i are then defined as

$$odds_i = \frac{p_i}{1 - p_i}$$

Taking the ***log*** of the odds removes the range restrictions.

$$log_e\left(\frac{p_i}{1-p_i}\right)=\sum_j \theta_j x_{ij} = X_i\theta$$

This way we map the probabilities from the [0; 1] range to the entire number line (real value).

Logistic Regression

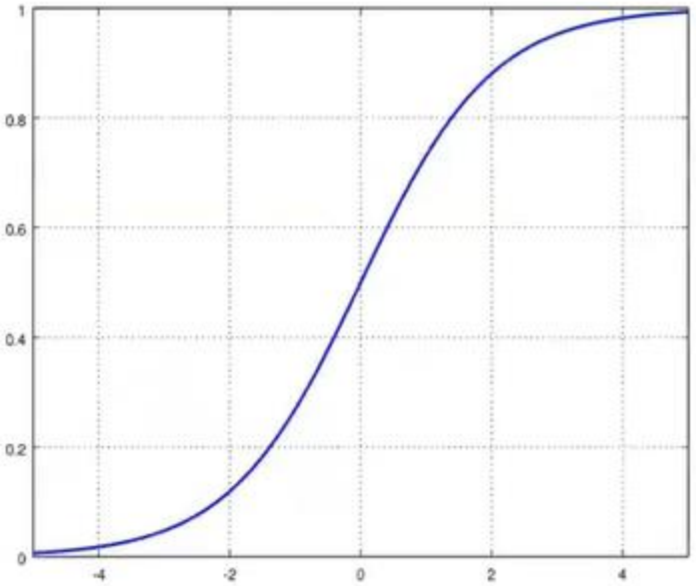
HYPOTHESIS FUNCTION

We need to calculate P_i

$$\log_e\left(\frac{p_i}{1-p_i}\right) = x_i\theta$$

$$\frac{p_i}{1-p_i} = e^{x_i\theta}$$

$$p_i = \frac{e^{x_i\theta}}{1+e^{x_i\theta}} = \frac{1}{1+e^{-x_i\theta}}$$



Standard logistic sigmoid function

$\log_2 8=3$

$\log_4 16=2$

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

$$p_i = g(\theta x) = \frac{1}{1+e^{-\theta x}}$$

LOGISTIC REGRESSION MODEL

Linear Regression

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

Logistic Regression

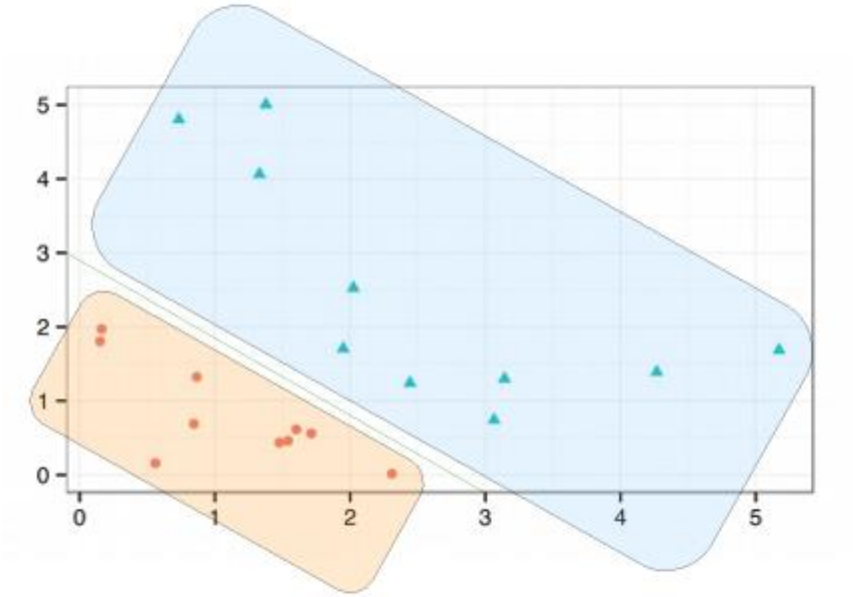
$$g(\theta^t x) = \begin{cases} 1, & \frac{1}{1+e^{-\theta x}} \geq 0.5 \\ 0, & 1 - \frac{1}{1+e^{-\theta x}} < 0.5 \end{cases}$$

$$p(y_i = 1 \mid x_i, \theta) = \frac{1}{1 + e^{-\theta^t x}}$$

$$p(y_i = 0 \mid x_i, \theta) = 1 - \frac{1}{1 + e^{-\theta^t x}}$$

$$p(y_i \mid x_i : \theta) = \left(\frac{1}{1 + e^{-\theta^t x}} \right)^{y_i} \left(1 - \frac{1}{1 + e^{-\theta^t x}} \right)^{1 - y_i}$$

Logistic Regression



- **If** $h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$

  and $\theta = \begin{bmatrix} -3 \\ 1 \\ 1 \end{bmatrix}$

- **Prediction y = 1 whenever**

$$\begin{aligned} \theta^T x &\geq 0 \\ \Leftrightarrow \quad -3 + x_1 + x_2 &\geq 0 \\ \Leftrightarrow \quad x_1 + x_2 &\geq 3 \end{aligned}$$

Logistic Regression

$$l(\theta) = \prod_{i=1}^{m} \left(\frac{1}{1+e^{-\theta^t x}} \right)^{y_i} \left(1 - \frac{1}{1+e^{-\theta^t x}} \right)^{1-y_i}$$

We can simplify L(ϴ) by taking its ***log*** and then differentiate to get the gradient.

$$j(\theta) = l(\theta) = \sum_{1}^{m} \left[y_i \log\left(\frac{1}{1+e^{-\theta^t x}} \right) + (1-y_i) \log\left(1 - \frac{1}{1+e^{-\theta^t x}} \right) \right].$$

$$\frac{d}{d\theta} j(\theta) = \frac{d}{d\theta} \sum_{1}^{m} \left[y_i \log\left(\frac{1}{1+e^{-\theta^t x}} \right) + (1-y_i) \log\left(1 - \frac{1}{1+e^{-\theta^t x}} \right) \right]$$

$$= \sum_{i=1}^{m} \left(y_i - \frac{1}{1+e^{-\theta^t x_i}} \right) x_i$$

Logistic Regression

Gradient Descent for logistic regression

- **Gradient Descent to minimize logistic regression cost function**

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

with identical algorithm as for linear regression

$$
\begin{aligned}
&\textbf{while}\ \ \text{not}\ \ \text{converged}:\\
&\quad \textbf{for}\ \ \text{all}\ \ j:\\
&\qquad tmp_j := \theta_j - \alpha\frac{1}{m}\sum_{i=1}^{m}\left(h_\theta(x^{(i)}) - y^{(i)}\right)x_j^{(i)}\\
&\quad \theta := \begin{bmatrix} tmp_0 \\ \vdots \\ tmp_n \end{bmatrix}
\end{aligned}
$$

Thank You