L11 Logistic Regression

Prof. Xun Jiao

Before Class

- Midterm/Test – Mar. 14
- HW3: Group project on predicting beer rating using scikit-learn
  - Find your team (2-3 people)
  - Write the python code, using scikit learning library
  - Predict beer rating based on
    - Features you select
    - ML methods you select
    - Parameters you select
  - Submit
    - A link of your Google Colab Python code (remember the sharing setting)
    - A report (1 page) document your features, ML method, and accuracy
  - Training/Testing Data
    - Use the train_test_split() in Scikit-Learn

Review

- Linear regression
- Gradient descent
- Feature Design

Review

- Machine learning
  - supervised learning
    - regression
    - classification
  - unsupervised learning

Outline

- Classification Overview
- Logistic regression
  - Hypothesis representation
  - Decision boundary
  - Cost function

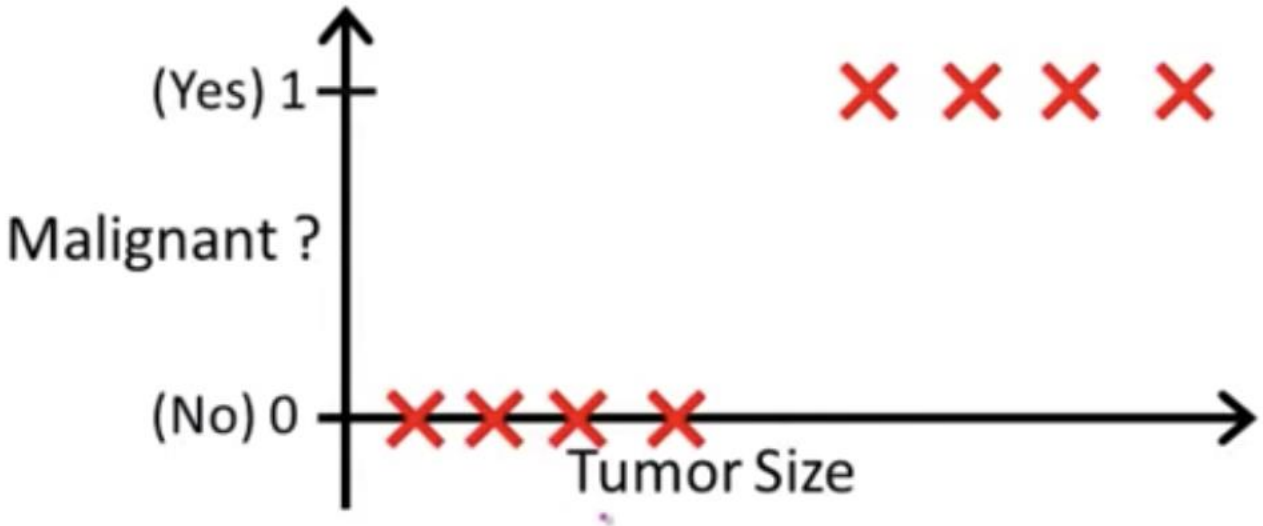
Classification

- Email: Spam/not spam
- Diagnose: sick/not sick
- Transaction: fraud/not fraud
- Image: cat/not cat
- Tumor: Malignant/benign

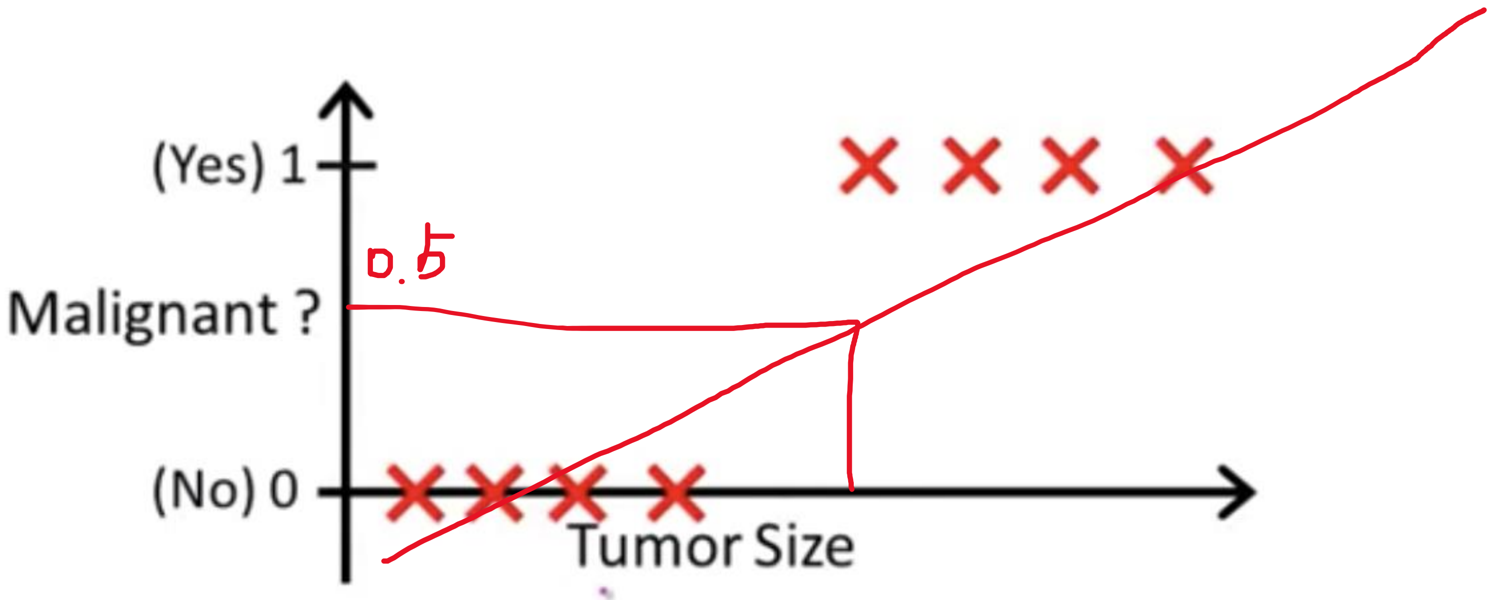
$$f(\text{data}) \overset{?}{\to} \text{labels}$$

{0,1}, {True, False}

{1, ... , N}



If we apply linear regression: $X\theta = y$



If we apply linear regression: $X\theta = y$

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"

If we apply linear regression: $X\theta = y$

Linear regression does not work well in classification problem.

Another problem

Classification: y = 0 or 1

$h_\theta(x)$ can be > 1 or < 0

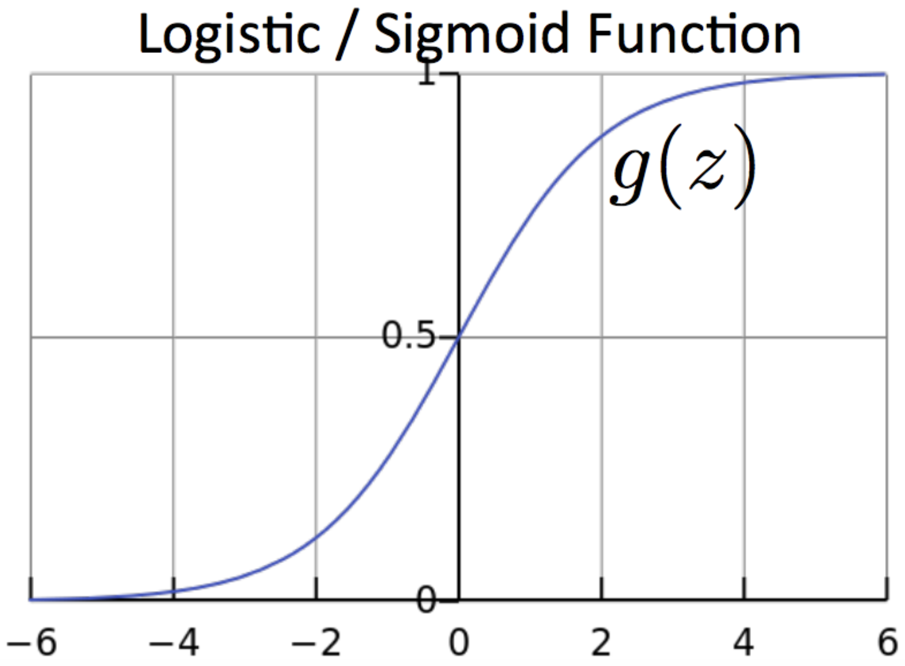
Logistic Regression: $0 \leq h_\theta(x) \leq 1$

Classification algorithm!

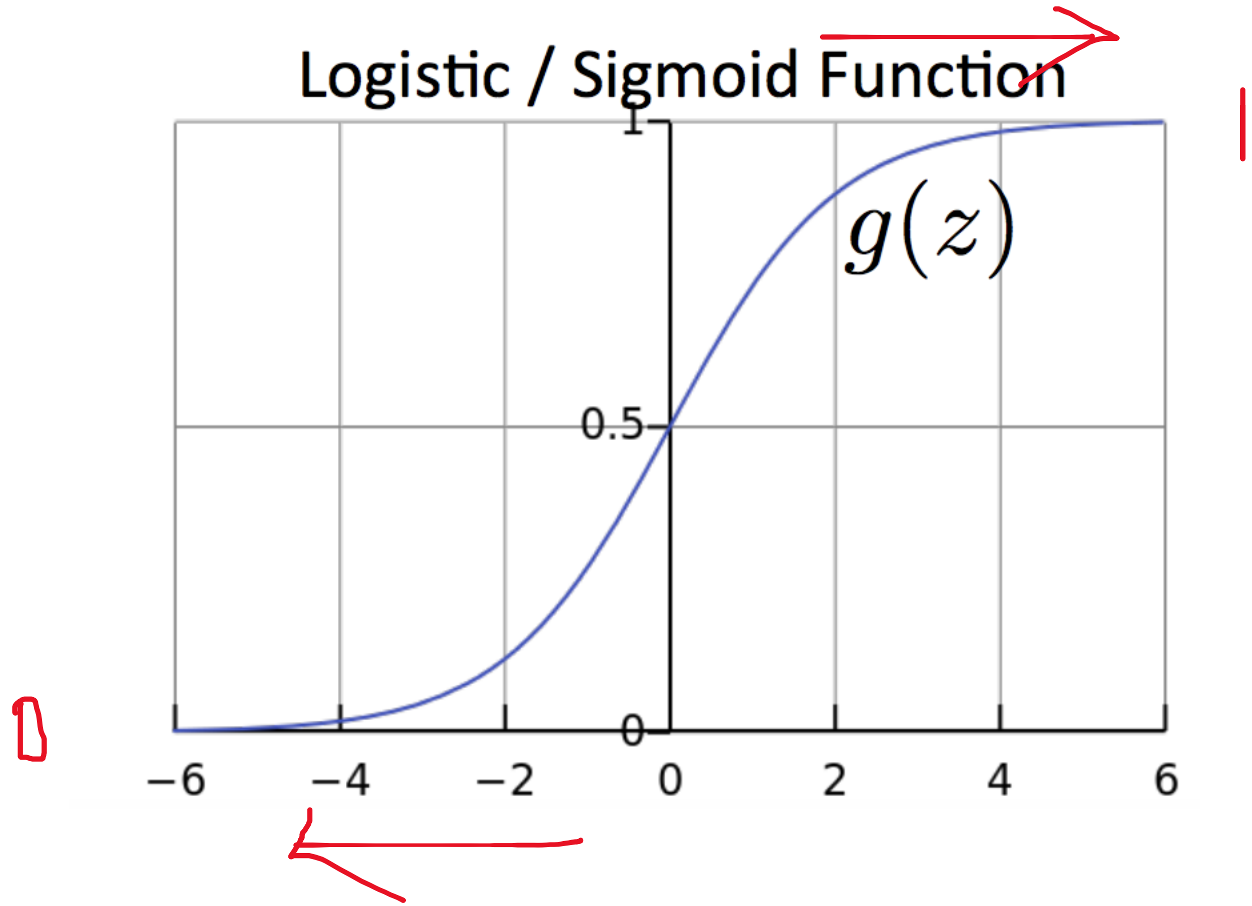
Hypothesis Representation

$$0 \leq h_{\boldsymbol{\theta}}(\boldsymbol{x}) \leq 1 \quad \begin{cases} h_{\boldsymbol{\theta}}(\boldsymbol{x}) = \boldsymbol{\theta}^\intercal \boldsymbol{x} \\ h_{\boldsymbol{\theta}}(\boldsymbol{x}) = g\left(\boldsymbol{\theta}^\intercal \boldsymbol{x}\right) \end{cases}$$

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



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


Logistic / Sigmoid Function
1
0.5
0
−6
−4
−2
0
2
4
6
$g(z)$
1
0

Interpretation of Logistic Regression

- The probability of positive prediction

$h_{\boldsymbol{\theta}}(\boldsymbol{x})$ = estimated $p(y = 1 \mid \boldsymbol{x}; \boldsymbol{\theta})$

Example: Cancer diagnosis from tumor size

$$\boldsymbol{x} = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} 1 \\ \text{tumorSize} \end{bmatrix}$$

$$h_{\boldsymbol{\theta}}(\boldsymbol{x}) = 0.7$$

→ Tell patient that 70% chance of tumor being malignant

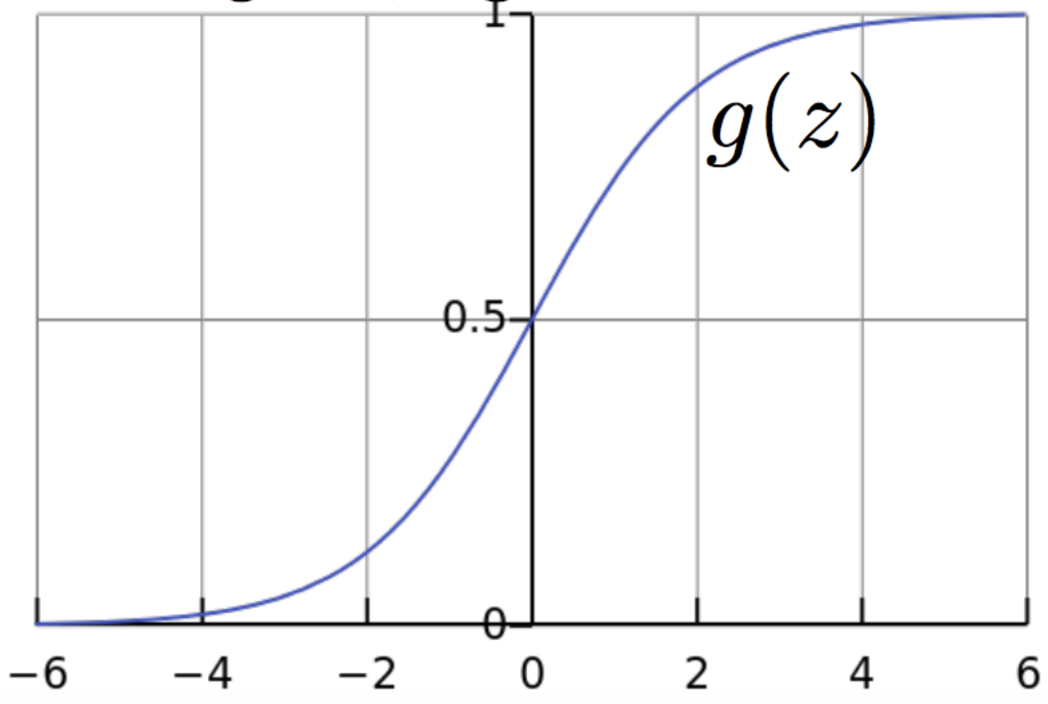
Note that: $p(y = 0 \mid \boldsymbol{x}; \boldsymbol{\theta}) + p(y = 1 \mid \boldsymbol{x}; \boldsymbol{\theta}) = 1$

Therefore, $p(y = 0 \mid \boldsymbol{x}; \boldsymbol{\theta}) = 1 - p(y = 1 \mid \boldsymbol{x}; \boldsymbol{\theta})$

Logistic / Sigmoid Function

$$h_{\boldsymbol{\theta}}(\boldsymbol{x}) = g\left(\boldsymbol{\theta}^{\intercal}\boldsymbol{x}\right)$$

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

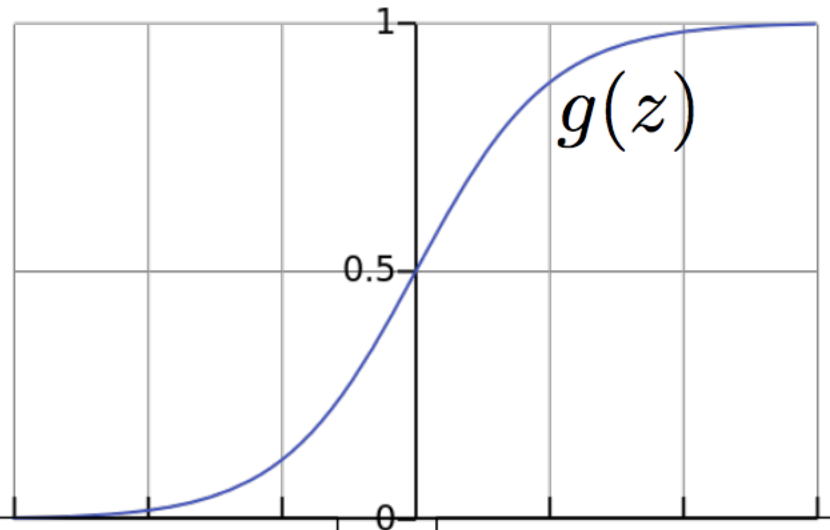


If g(z) >= 0.5, predict y = ? → $\boldsymbol{\theta}^{\intercal}\boldsymbol{x}$ = ??

If g(z) < 0.5, predict y = ? → $\boldsymbol{\theta}^{\intercal}\boldsymbol{x}$ = ??

$$h_{\boldsymbol{\theta}}(\boldsymbol{x}) = g\left(\boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{x}\right)$$

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

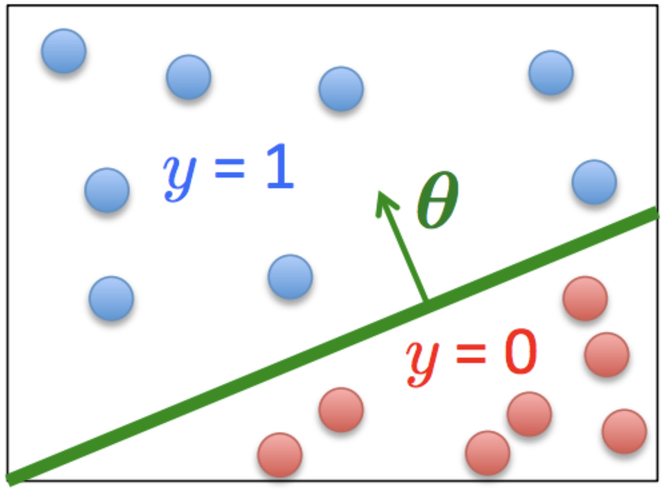


$\boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{x}$ should be large <u>negative</u> values for negative instances

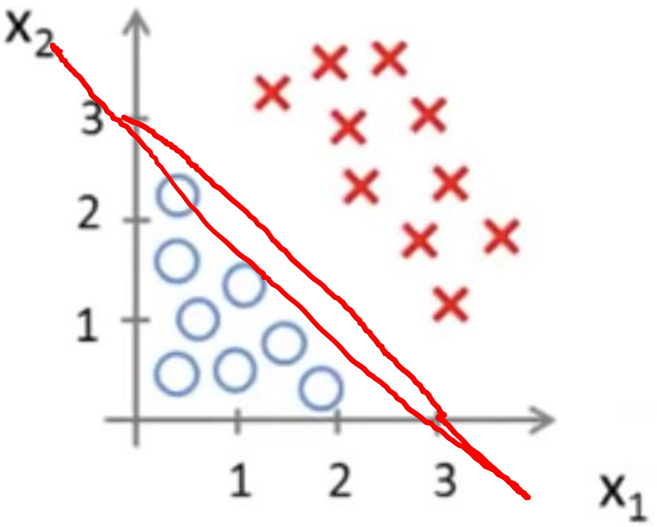
$\boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{x}$ should be large <u>positive</u> values for positive instances

Assume a threshold and...

- Predict y = 1 if $h_{\boldsymbol{\theta}}(\boldsymbol{x}) \geq 0.5$
- Predict y = 0 if $h_{\boldsymbol{\theta}}(\boldsymbol{x}) < 0.5$



Decision Boundary



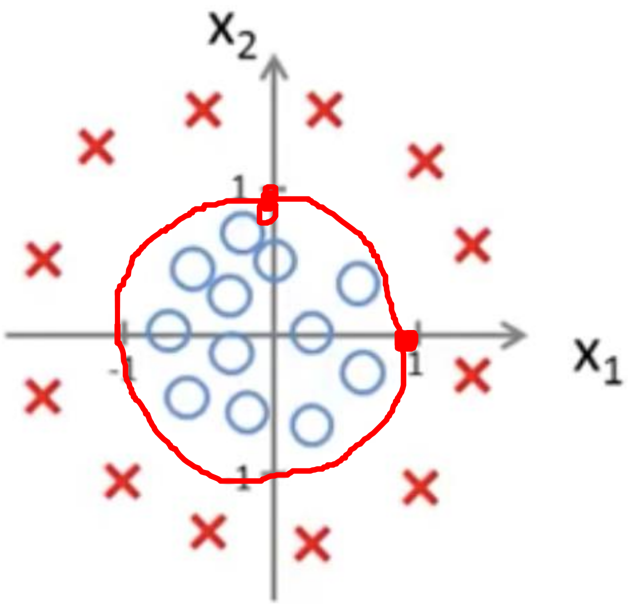
$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$$

$-3 \qquad 1 \qquad 1$

Predict "$y = 1$" if $-3 + x_1 + x_2 \geq 0$

Draw a line (decision boundary).

Non-linear decision boundaries



$$h_\theta(x) = g(\overset{-1}{\theta_0} + \overset{0}{\theta_1} x_1 + \overset{0}{\theta_2} x_2 + \underset{1}{\theta_3} x_1^2 + \underset{1}{\theta_4} x_2^2)$$

Derive the decision boundary for me?

How do we find theta?

- Given $\left\{\left(\boldsymbol{x}^{(1)}, y^{(1)}\right), \left(\boldsymbol{x}^{(2)}, y^{(2)}\right), \ldots, \left(\boldsymbol{x}^{(n)}, y^{(n)}\right)\right\}$
  where $\boldsymbol{x}^{(i)} \in \mathbb{R}^d,\ y^{(i)} \in \{0, 1\}$

- Model: $h_{\boldsymbol{\theta}}(\boldsymbol{x}) = g\left(\boldsymbol{\theta}^\mathsf{T}\boldsymbol{x}\right)$

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

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_d \end{bmatrix} \qquad \boldsymbol{x}^\mathsf{T} = \begin{bmatrix} 1 & x_1 & \ldots & x_d \end{bmatrix}$$