Artificial Intelligence

Lecture 11. Logistic Regression

Spring 2022

Prof. Jonghoon Chun, Ph.D.

E-mail: jchun@mju.ac.kr

Lecture Note: http://lms.mju.ac.kr

Agenda

- Classification concept review
- Linear Regression
- Logistic Regression

LOGISTIC REGRESSION

Classification Revisit

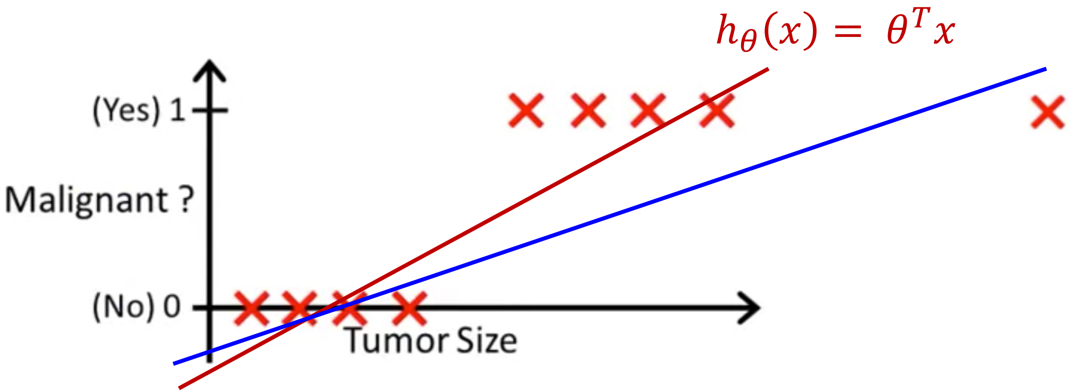
- Classification
  - Email: Spam / Not Spam?
  - Online Transactions: Fraudulent (Yes / No)?
  - Tumor: Malignant / Benign?

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

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

$y \in \{0, 1, 2, \ldots, n\}$ → multiclass problem

Problems with Linear Regression



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

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

Classification: y = 0 or 1

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

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

Logistic Regression is a "**Classification**" algorithm!

Hypothesis Representation

- Hypothesis Representation
  - What is the function we're going to use to represent our hypothesis when we have a classification problem?

- Logistic Regression Model

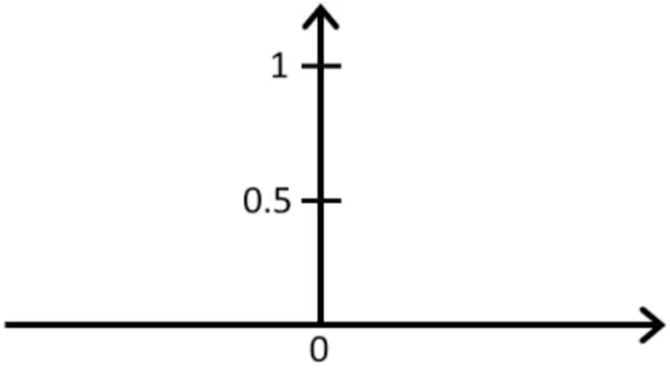
$$0 \le h_\theta(x) \le 1$$

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

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

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

Sigmoid function
Logistic function



Interpretation of Hypothesis Output

- $h_\theta(x)$ = estimated probability that y = 1 on input x

- Example:

  If $x = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = \begin{bmatrix} 1 \\ TumorSize \end{bmatrix}$

  $h_\theta(x)$ = 0.7

  Tell patient that 70% chance of tumor being malignant

$$h_\theta(x) = P(y = 1 \mid x ; \theta)$$

"Probability that y = 1, given x, parameterized by θ"

$$y = 0 \text{ or } 1$$

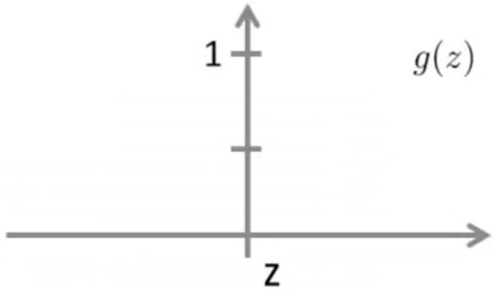
$$P(y = 0 \mid x ; \theta) + P(y = 1 \mid x ; \theta) = 1$$
$$P(y = 0 \mid x ; \theta) = 1 - P(y = 1 \mid x ; \theta)$$

DECISION BOUNDARY

Logistic Regression Recap

$$h_\theta(x) = g(\theta^T x) = P(y = 1 \mid x\,;\theta)$$

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



Suppose predict "$y = 1$" if $h_\theta(x) \geq 0.5$

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

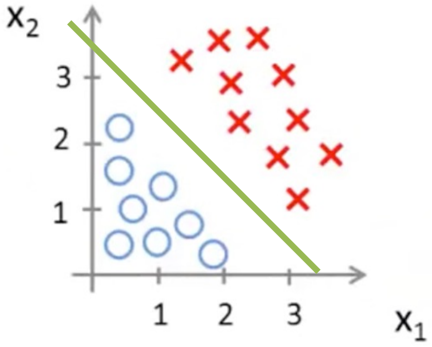
predict "$y = 0$" if $h_\theta(x) < 0.5$

$$\theta^T x < 0$$

$g(z) \geq 0.5$
when $z \geq 0$

$h_\theta(x) = g(\theta^T x) \geq 0.5$
whenever $\theta^T x \geq 0$

Decision Boundary



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

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

θ_0 = -3, θ_1 = 1, θ_2 = 1

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

$\theta^T x$

$x_1 + x_2 \geq 3$

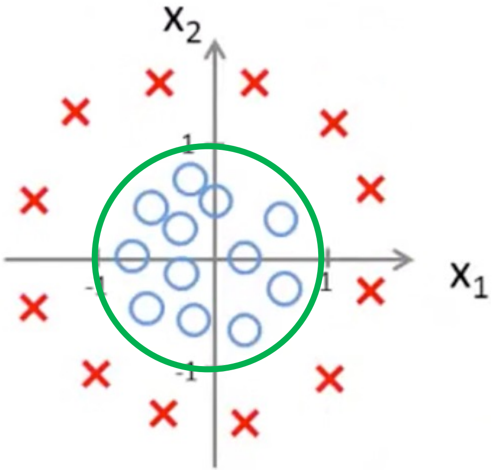
$x_1 + x_2 = 3$

$h_\theta(x) = 0.5$

$x_1 + x_2 < 3$

$y = 0$

Non-linear Decision Boundaries



$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2)$$

$$\theta = \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

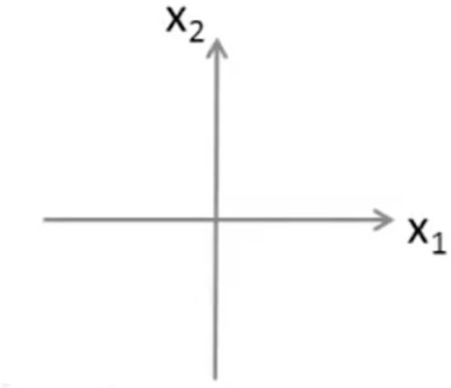
$\theta_0 = -1,\ \theta_1 = 0,\ \theta_2 = 0,\ \theta_3 = 1,\ \theta_4 = 1$

Predict "$y = 1$" if $-1 + x_1^2 + x_2^2 \geq 0$

$x_1^2 + x_2^2 = 1$ $\quad$ $x_1^2 + x_2^2 \geq 1$

The training set may be used to fit the parameters θ.
But, once you have the parameters θ,
that is what defines the decisions boundary!

More complex ones...



$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_1^2 x_2 + \theta_5 x_1^2 x_2^2 + \theta_6 x_1^3 x_2 + \dots)$$

COST FUNCTION

How to fit the parameters

- Optimization Objectives

- Cost function
  - Use to fit the parameters

Problem Recap

Training set: $\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \cdots, (x^{(m)}, y^{(m)})\}$

m examples $\quad x \in \begin{bmatrix} x_0 \\ x_1 \\ \dots \\ x_n \end{bmatrix} \quad x_0 = 1, y \in \{0, 1\}$

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

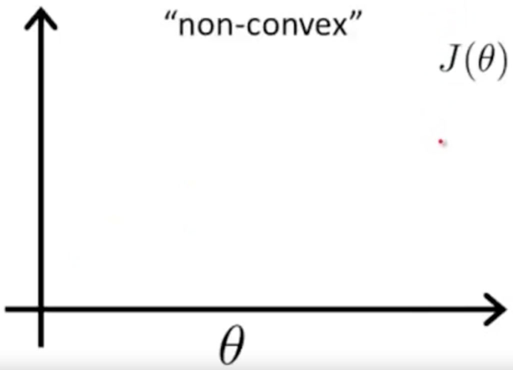
How to choose parameters θ?

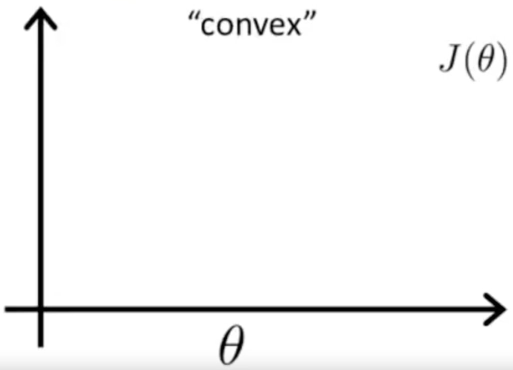
Cost Function

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

$Cost(h_\theta(x^{(i)}), y^{(i)})$

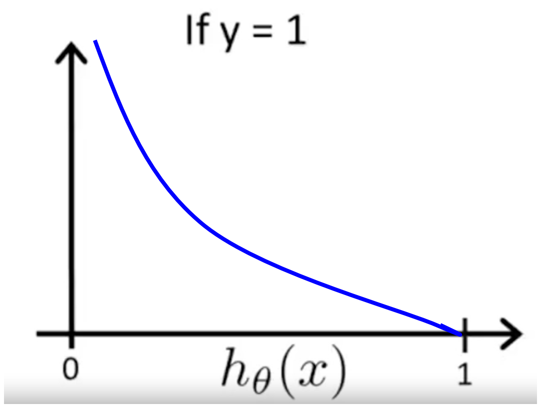
$$\mathrm{Cost}(h_\theta(x^{(i)}), y^{(i)}) = \frac{1}{2}\left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$





Logistic Regression Cost Function

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



$\text{Cost} = 0 \text{ if } y = 1, h_\theta(x) = 1$

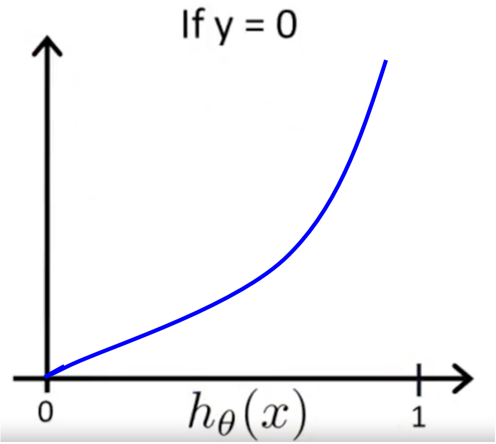
But as $h_\theta(x) \rightarrow 0$

$Cost \rightarrow \infty$

Captures intuition that if $h_\theta(x) = 0$, (predict $P(y = 1|x;\theta) = 0$), but $y = 1$, we'll penalize learning algorithm by a very large cost.

Logistic Regression Cost Function

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



Note that above cost function guarantees
that J(θ) is “convex” for logistic regression!

ACKNOWLEDGMENTS

- Many of the slides and examples are borrowed from the open course by Coursera “**Machine Learning**” by A. Ng which were modified into their current form.

END