투빅스 12기 정규과정

ToBig's 11기 이영전

# Logistic Regression & Classification

Contents

## Unit 01 | Machine Learning

### ◆ Basic concept of ML

- What is ML?
- What is learning?
  - Supervised?
  - Unsupervised?
- What is regression?
- What is classification?

## ◆ About ML

세기의 대국. 2016年 3月 → 역사적인 순간...

# ◆ About ML

Machine Learning. Super Power

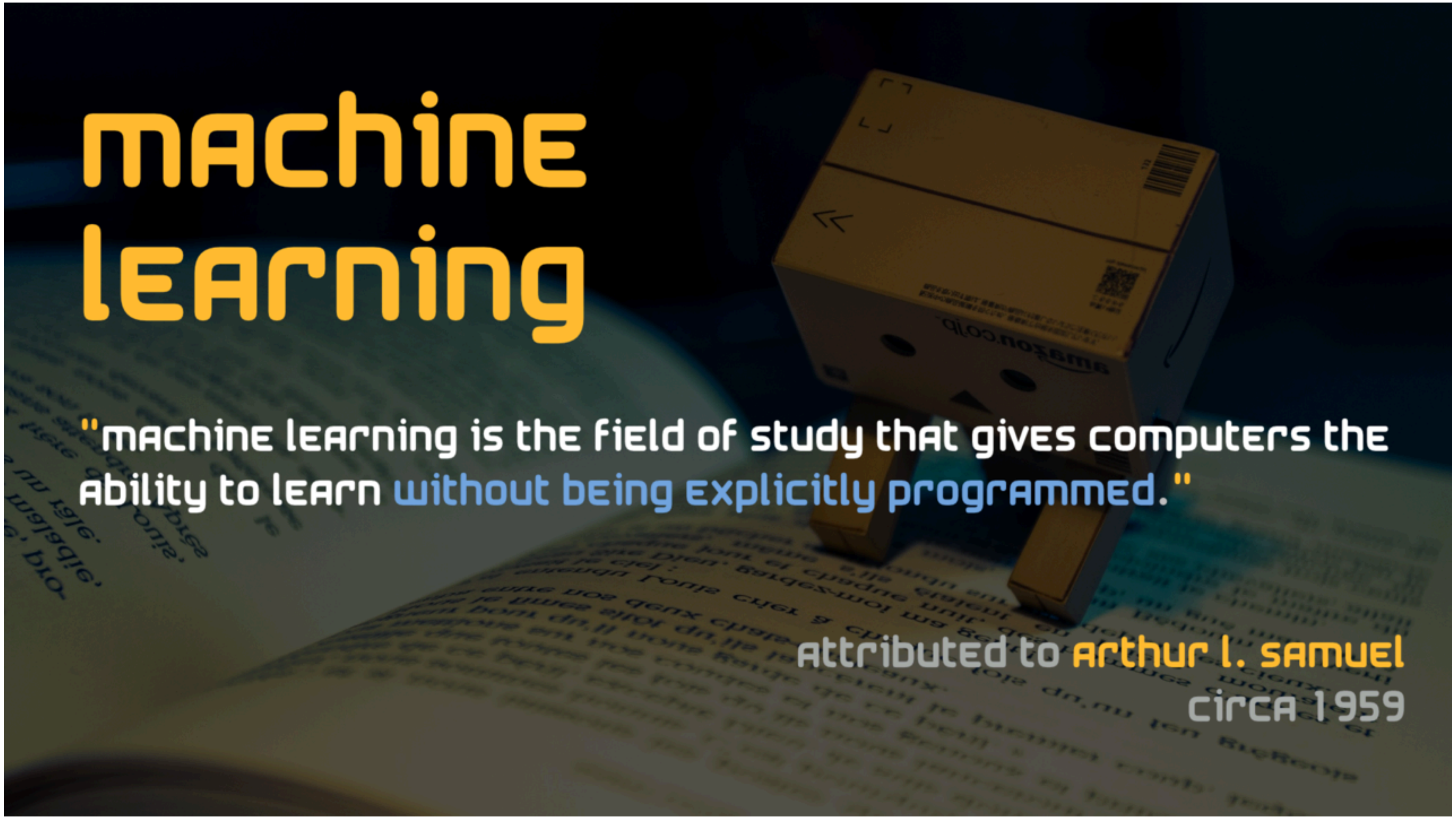

GPU Technology Conference 2015 - Dr. Andrew Ng

## ◆ About ML

# Machine Learning

- Limitations of explicit programming ←손 코딩의 한계 (Logic Coding)
  - Spam filter: many rules
  - Automatic driving: too many rules

## ◆ About ML


machine learning
"machine learning is the field of study that gives computers the ability to learn without being explicitly programmed."
attributed to arthur l. samuel
circa 1959

## ◆ 머신러닝의 종류 – Label의 유무에 따른 구분

- 1. Supervised Learning

  ✓ **learning with labeled examples** – training set

  ✓ **Label이 존재하는 데이터**를 이용해 독립변수 X와 종속변수 Y사이의 관계를 찾아 Test 데이터의 Y를 예측하기 위한 학습

  ✓ 분류, 회귀 모델이 Supervised Learning에 해당됨

| | X_1 | X_2 | … | X_d |
|---|---|---|---|---|
| Data_1 | … | … | … | … |
| Data_2 | … | … | … | … |
| … | … | … | … | … |
| Data_n | … | … | … | … |

Y=f(x)

| Y |
|---|
| … |
| … |
| … |
| … |

# ◆ 머신러닝의 종류 – Label의 유무에 따른 구분

- 2. Unsupervised Learning

  ✓ **Un-labeled data**

  ✓ **Label이 존재하지 않는 데이터**를 이용해 데이터에 내재된 특징을 분석하기 위한 학습

  ✓ 군집화, 연관규칙분석이 비지도학습에 해당됨

| | X_1 | X_2 | … | X_d |
|---|---|---|---|---|
| Data_1 | … | … | … | … |
| Data_2 | … | … | … | … |
| … | … | … | … | … |
| Data_n | … | … | … | … |

# ◆ 머신러닝의 종류 – 학습목적에 따른 구분

## Types of supervised learning

크게 세 가지의 방법이 존재한다.

① • Predicting final exam score based on time spent
  - regression

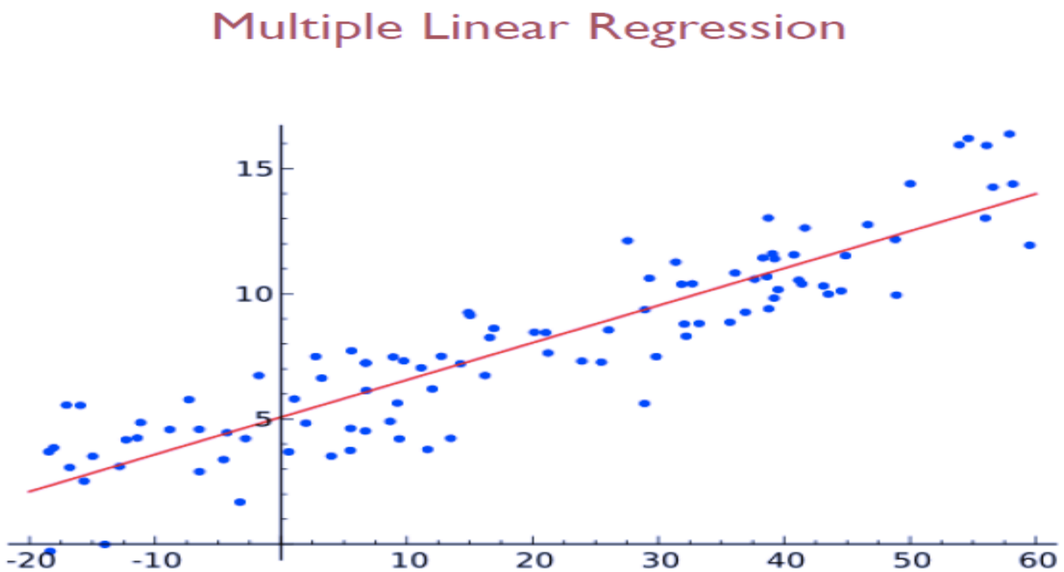
② • Pass/non-pass based on time spent
  - binary classification // 2가지 문제 분류.

③ • Letter grade (A, B, C, E and F) based on time spent
  - multi-label classification

# ◆ 머신러닝의 종류 – 학습목적에 따른 구분

- 1. Regression

  ✓ 연속형(Continuous)변수를 예측하는 방법론
  ✓ 다중선형회귀분석, 인공신경망, SVR(Support Vector Regression) 등

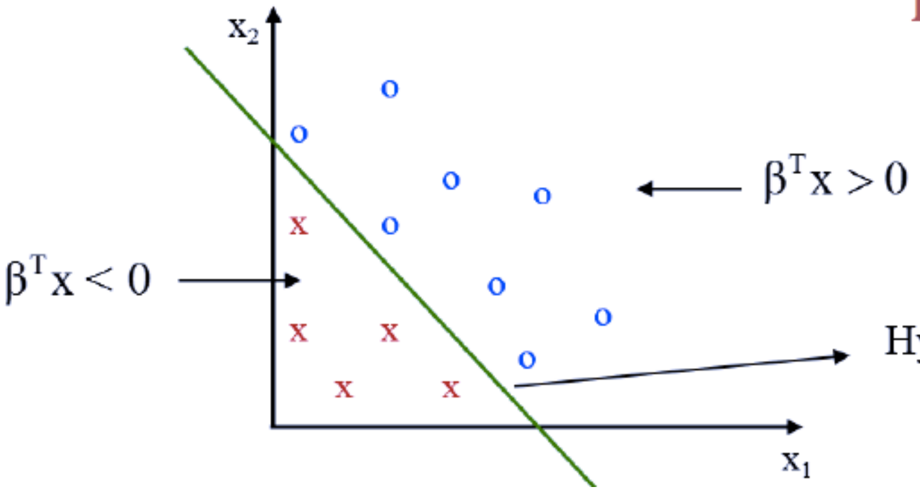


# ◆ 머신러닝의 종류 – 학습목적에 따른 구분

- 2. Classification

  - ✓ 명목형(Categorical) 변수를 예측하는 방법론
  - ✓ Logistic Regression, Naïve Bayes, Decision Tree, SVM 등

## Logistic Regression




Hyperplane: $\beta^T x = 0$ where $\beta = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}$

**Classifier**

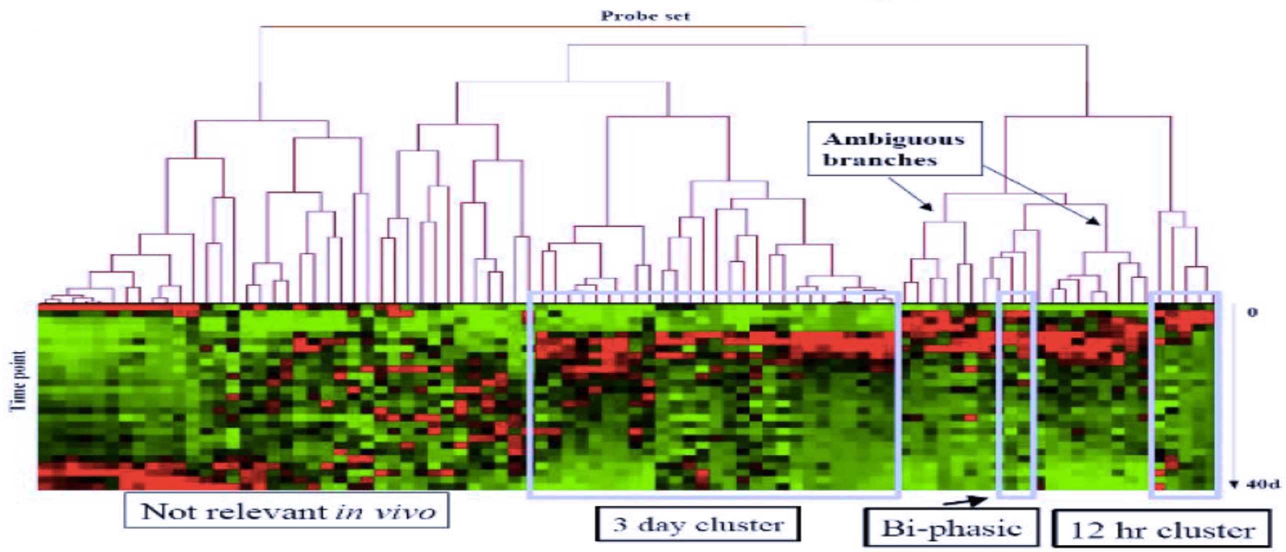
$$y = \frac{1}{\left(1+\exp(-\beta^T x)\right)} \qquad \begin{pmatrix} y \to 1 & \text{if} & \beta^T x \to \infty \\ y = \frac{1}{2} & \text{if} & \beta^T x = 0 \\ y \to 0 & \text{if} & \beta^T x \to -\infty \end{pmatrix}$$

# ◆ 머신러닝의 종류 – 학습목적에 따른 구분

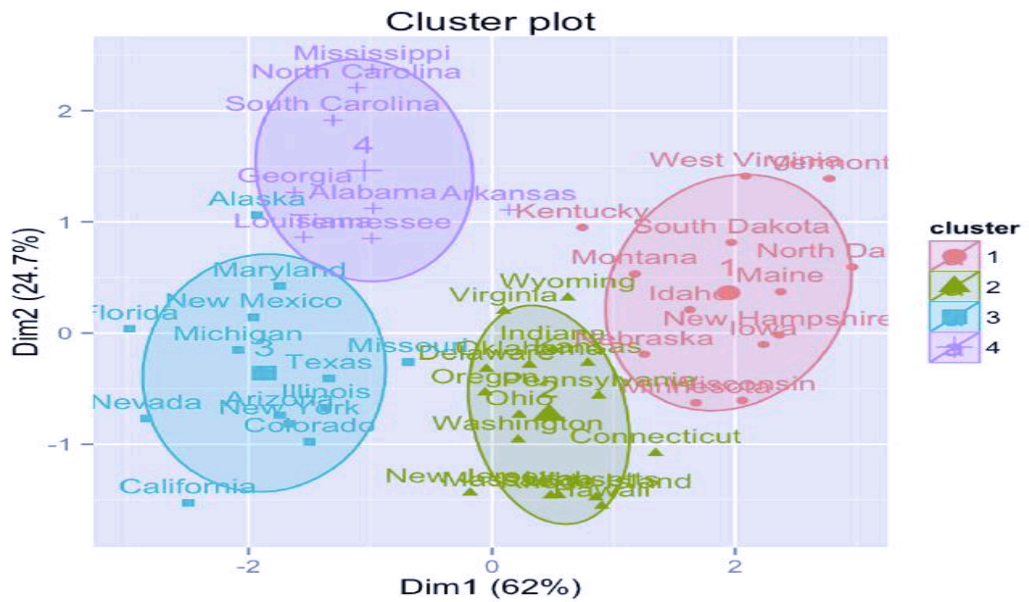
- 3. Clustering

  ✓ 유사한 개체들의 집단을 판별하는 방법론
  ✓ Hierarchical Clustering, K-Means Clustering 등

Hierarchical Clustering




K-Means Clustering



Unit 01 | Machine Learning

## ◆ 머신러닝 학습 목표

- 우리가 회귀 분석을 배우는 이유

**오늘 배우는 Logistic Regression 어디에 쓰이나요?**

**왜 배우죠?**

## ◆ 머신러닝 학습 목표

# Goals

Deep Learning을 이해하기 위한 초석.

- Basic understanding of machine learning algorithms
  - Linear regression, Logistic regression (classification)
  - Neural networks, Convolutional Neural Network, Recurrent Neural Network
- Solve your problems using machine learning tools
  - Tensorflow and Python

Unit 01 | Machine Learning

◆ 머신러닝 학습 목표

Goals

Deep Learning을 이해하기 위한 초석

- Basic understanding of machine learning algorithms
  - Linear regression, Logistic regression (classification)
  - Neural networks, Convolutional Neural Network, Recurrent Neural Network
- Solve your problems using machine learning tools
  - Tensorflow and Python

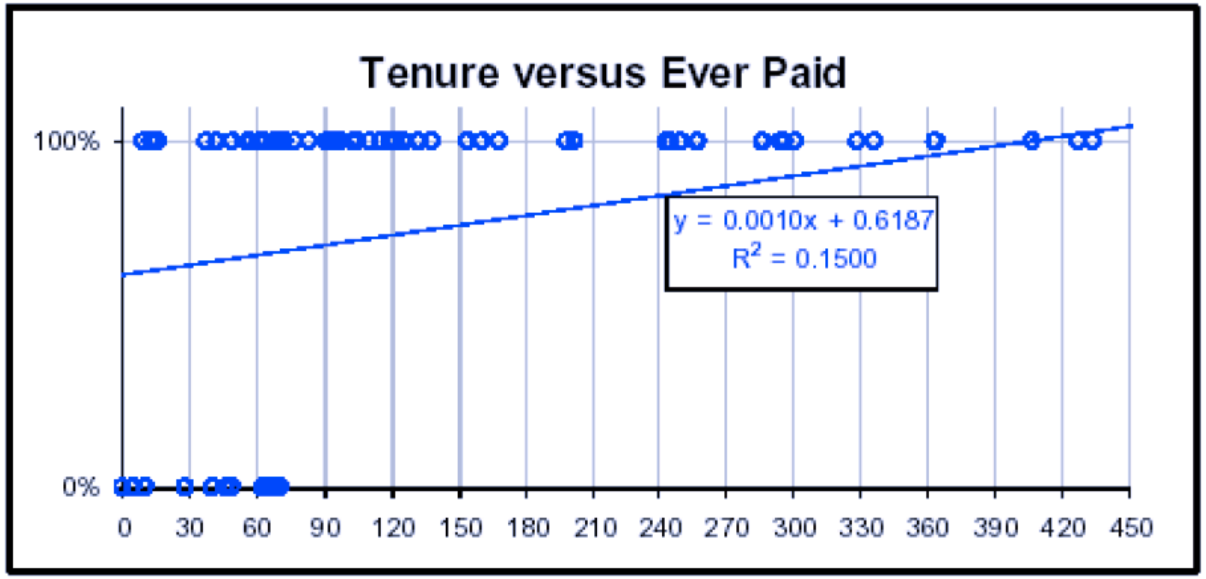
**Machine Learning의 기본을 잘 다져놓으면**
**Deep Learning을 학습 할 때 큰 도움이 된다!**

# Contents

## ◆ Logistic Regression

- Logistic Regression의 목적
  - ✓ Binary한 종속변수에 대하여 **회귀식의 형태로 모형을 추정**함
  - ✓ Machine Learning의 종류 중 Supervised Learning, Classification에 속함

$$P(Y=1)=\beta_0+\beta_1 x_1+\beta_2 x_2+\ldots+\beta_p x_p+\varepsilon$$



## ◆ Logistic Regression


**왜 배우는지**

**무엇을 배우는지**

**어떻게 사용하는지**

**어떻게 응용할 수 있는지**

**질문의 꼬리를 물며 배워보자!**

## Unit 02 | Logistic Regression

### ◆ Logistic Regression I

- 로지스틱 회귀모델 배경
- 로지스틱 회귀모델 형태
- Odds

### ◆ Logistic Regression II

- 파라미터 추정
- 로지스틱 회귀모델 결과 및 해석
- 로지스틱 회귀모델 예제

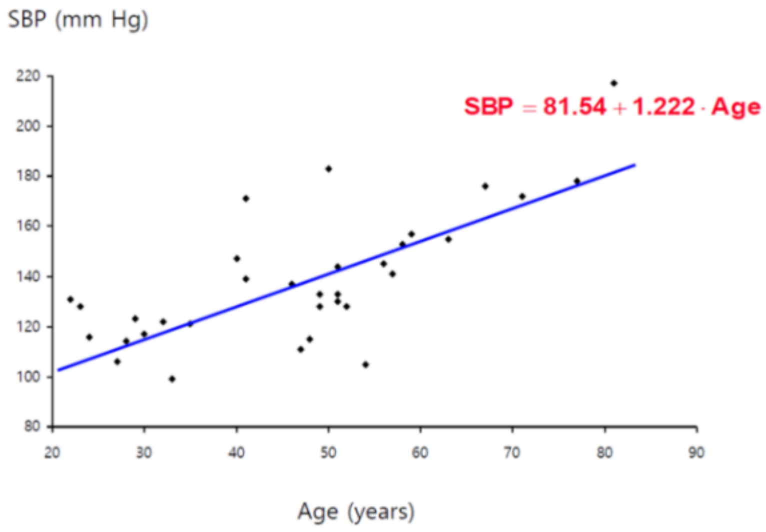
# ◆ Logistic Regression I

## • 0. 문제의식

### ✓ 다중선형회귀(Multiple Linear Regression)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \varepsilon$$

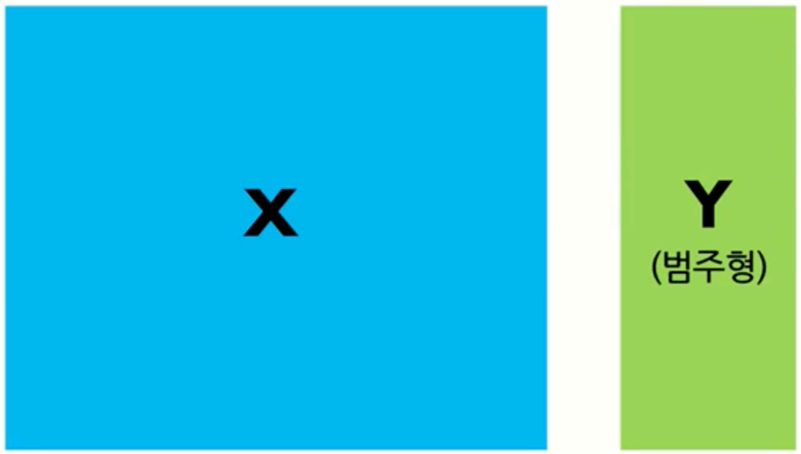
| Age | SBP | Age | SBP | Age | SBP |
|---|---|---|---|---|---|
| 22 | 131 | 41 | 139 | 52 | 128 |
| 23 | 128 | 41 | 171 | 54 | 105 |
| 24 | 116 | 46 | 137 | 56 | 145 |
| 27 | 106 | 47 | 111 | 57 | 141 |
| 28 | 114 | 48 | 115 | 58 | 153 |
| 29 | 123 | 49 | 133 | 59 | 157 |
| 30 | 117 | 49 | 128 | 63 | 155 |
| 32 | 122 | 50 | 183 | 67 | 176 |
| 33 | 99 | 51 | 130 | 71 | 172 |
| 35 | 121 | 51 | 133 | 77 | 178 |
| 40 | 147 | 51 | 144 | 81 | 217 |



# ◆ Logistic Regression I

- **0. 문제의식**

✓ **로지스틱 회귀모델 필요성**




- 범주형 반응변수
  - 이진변수 (반응변수 값 ∈ 0 or 1)
  - 멀티변수 (반응변수 값 ∈ 1 or 2 or 3이상)

- **선형회귀모델과는 다른 방식**으로 접근해야 될 필요성

# ◆ Logistic Regression I

## • 0. 문제의식

✓ **응용예제**

- 새로운 관측치가 왔을 때 이를 기존 범주 중 하나로 예측(범주예측)

✓ **로지스틱 회귀모델의 사용**

- 제품이 불량인지 양품인지 분류
- 고객이 이탈고객인지 잔류고객인지 분류
- 카드 거래가 정상인지 사기인지 분류
- 내원 고객이 질병이 있는지 없는지 분류
- …

Unit 02 | Logistic Regression

# ◆ Logistic Regression I

- 로지스틱 회귀모델 이론 배경

✓ **어떤 이론적 배경**을 가지고 출발했을까?

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad Y_i = 0 \text{ or } 1$$

Assume $E(\varepsilon_i) = 0$ / $E(Y_i) = \beta_0 + \beta_1 X_i$

Consider $Y_i$ to be a Bernoulli random variable

$$P(Y_i = 1) = \pi_i$$

$$P(Y_i = 0) = 1 - \pi_i$$

$$E(Y_i) = 1 \cdot \pi_i + 0 \cdot (1 - \pi_i) = \pi_i$$

$$E(Y_i) = \beta_0 + \beta_1 X_i = \pi_i$$

X값이 주어졌을 때 출력변수 Y가 1의 값을 가질 확률

## ◆ Logistic Regression

✓ 로지스틱 회귀모델 이론 배경

X 값이 주어졌을 때, Y가 특정범주 값을 가질 확률 값으로 표현되는 모델이다.

Consider $Y_i$ to be a Bernoulli random variable

$$P(Y_i = 1) = \pi_i$$

그럼, **<u>과연 이렇게 직선으로 표현하는 게 맞는 걸까?</u>**

$$E(Y_i) = 1 \cdot \pi_i + 0 \cdot (1 - \pi_i) = \pi_i$$

$$E(Y_i) = \beta_0 + \beta_1 X_i = \pi_i$$

X값이 주어졌을 때 출력변수 Y가 1의 값을 가질 확률

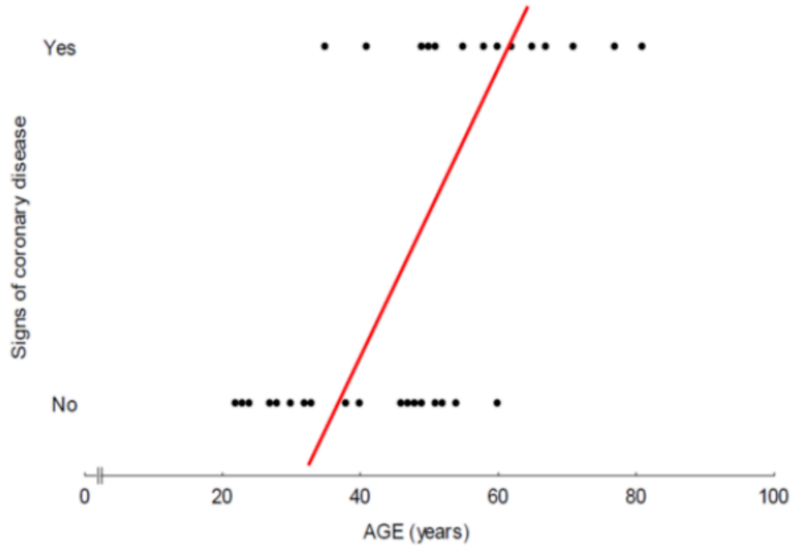
# ◆ Logistic Regression I

- 로지스틱 회귀모델

✓ 출력변수가 연속형이 아닌 이진범주형 질병유무?
✓ 두 변수 사이의 관계식은 선형인가?

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \varepsilon$$

| Age | CD | Age | CD | Age | CD |
|---|---|---|---|---|---|
| 22 | 0 | 40 | 0 | 54 | 0 |
| 23 | 0 | 41 | 1 | 55 | 1 |
| 24 | 0 | 46 | 0 | 58 | 1 |
| 27 | 0 | 47 | 0 | 60 | 1 |
| 28 | 0 | 48 | 0 | 60 | 0 |
| 30 | 0 | 49 | 1 | 62 | 1 |
| 30 | 0 | 49 | 0 | 65 | 1 |
| 32 | 0 | 50 | 1 | 67 | 1 |
| 33 | 0 | 51 | 0 | 71 | 1 |
| 35 | 1 | 51 | 1 | 77 | 1 |
| 38 | 0 | 52 | 0 | 81 | 1 |

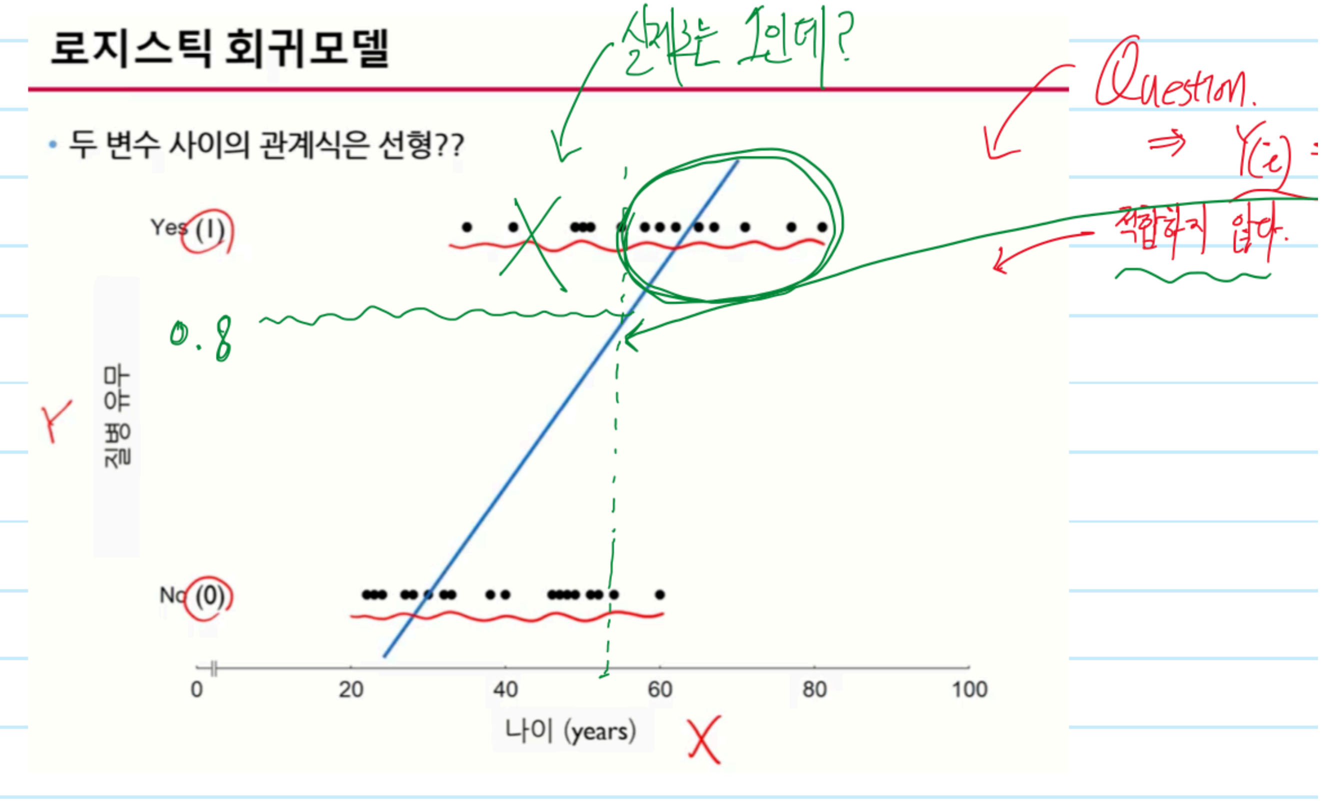


Unit 02 | Logistic Regression

## ◆ Logistic Regression I

- 로지스틱 회귀모델

✓ 두 변수 사이의 관계식을 직선으로 표현하는 게 맞을까?

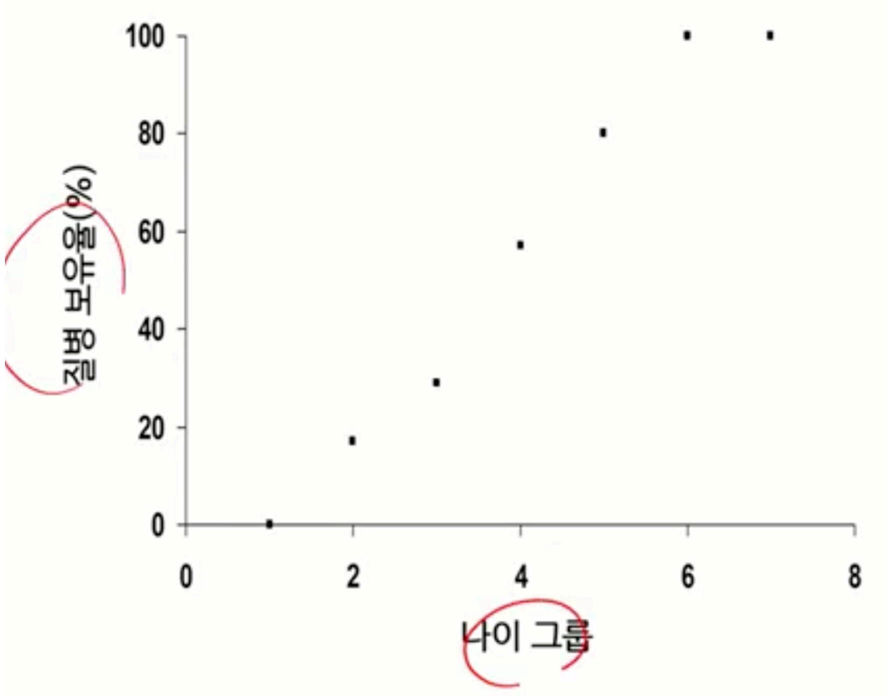


## ◆ Logistic Regression I

- 로지스틱 회귀모델

  ✓ 확률변수 – 나이그룹
  ✓ 각 나이 대별로 질병에 걸릴 확률

| | | 질병 | |
|---|---|---|---|
| 나이 그룹 | 그룹내 수 | 질병보유자 수 | % |
| 20 - 29 | 5 | 0 | 0 |
| 30 - 39 | 6 | 1 | 17 |
| 40 - 49 | 7 | 2 | 29 |
| 50 - 59 | 7 | 4 | 57 |
| 60 - 69 | 5 | 4 | 80 |
| 70 - 79 | 2 | 2 | 100 |
| 80 - 89 | 1 | 1 | 100 |



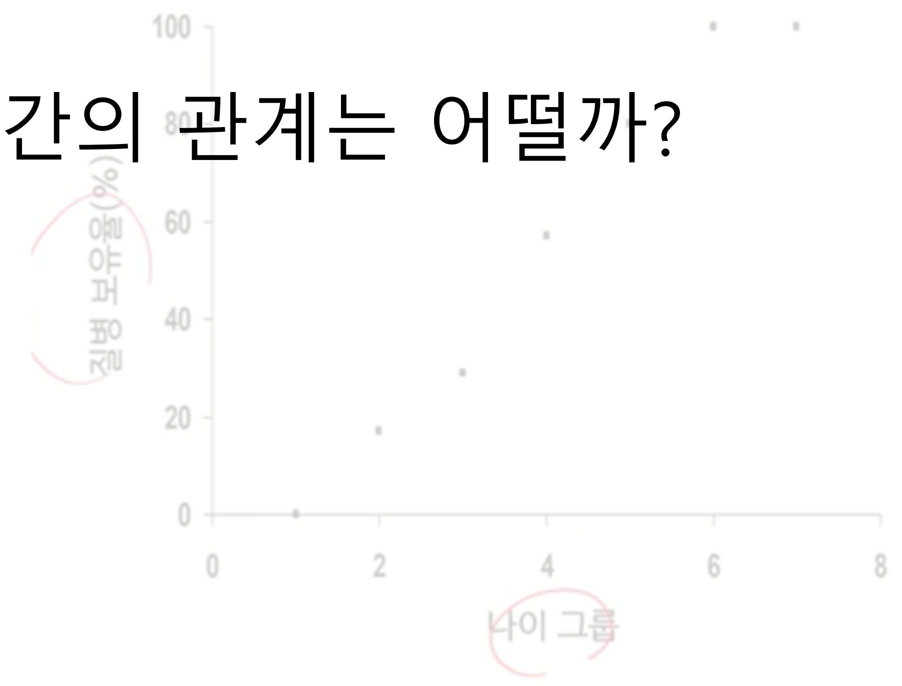
Unit 02 | Logistic Regression

## ◆ Logistic Regression I

- 로지스틱 회귀모델

✓ 확률변수 – 나이그룹
✓ 각 나이 대별로 질병에 걸릴 확률

| | | 질병 | |
|---|---|---|---|
| 나이 그룹 | 그룹내 수 | 질병보유자 수 | % |
| 20 - 29 | 5 | 0 | [illegible] |
| 30 - 39 | 6 | 1 | 17 |
| 40 - 49 | 7 | 2 | 29 |
| 50 - 59 | 7 | 4 | 57 |
| 60 - 69 | 5 | 4 | 80 |
| 70 - 79 | 2 | 2 | 100 |
| 80 - 89 | 1 | 1 | 100 |

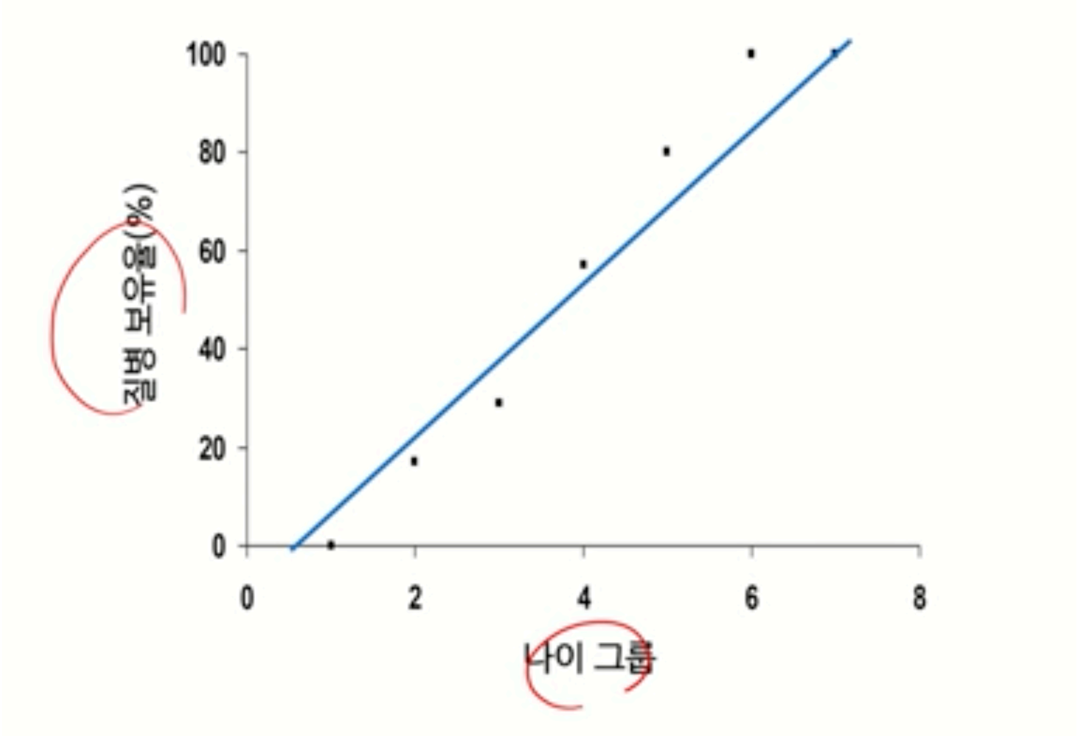


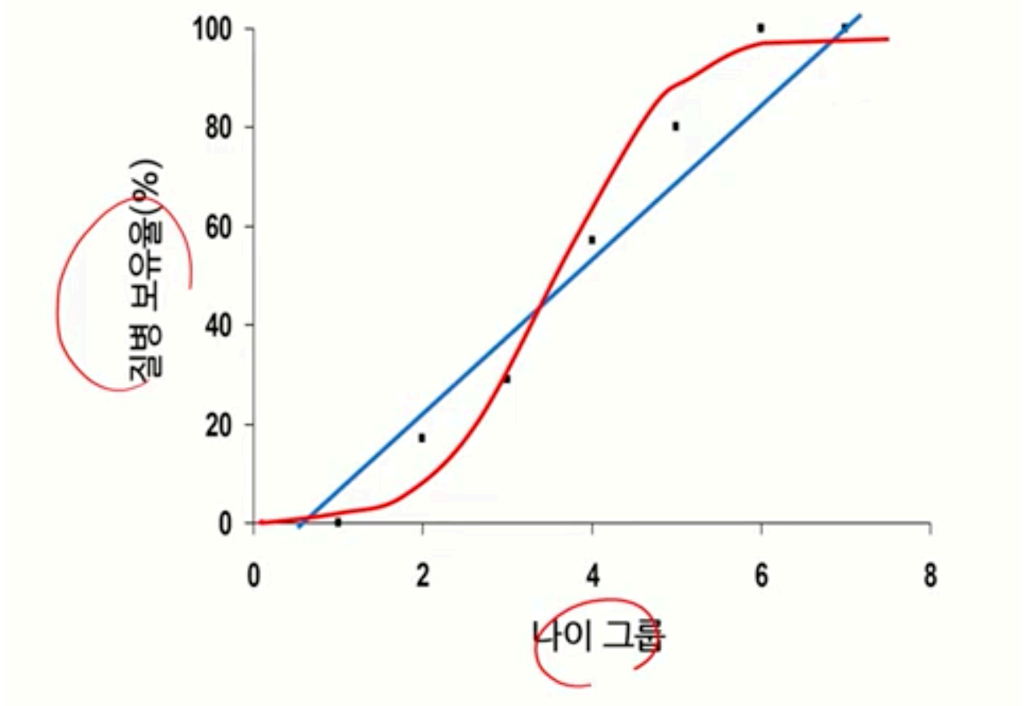

Q. 나이 그룹과 질병 보유율 간의 관계는 어떨까?

# ◆ Logistic Regression I

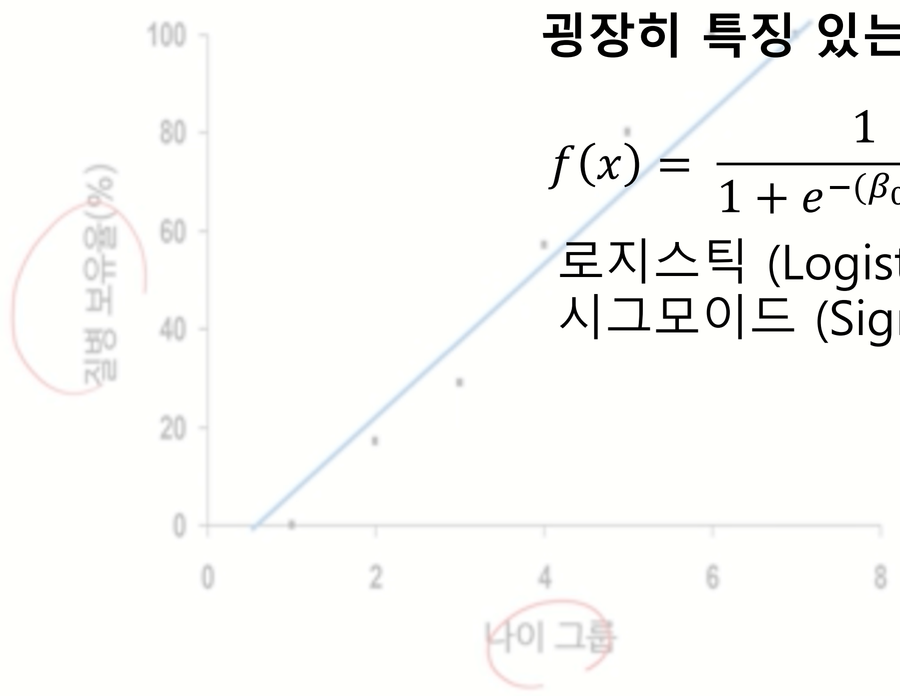
- 로지스틱 회귀모델

  ✓ 선형으로 표현
  ✓ S-Curve로 표현



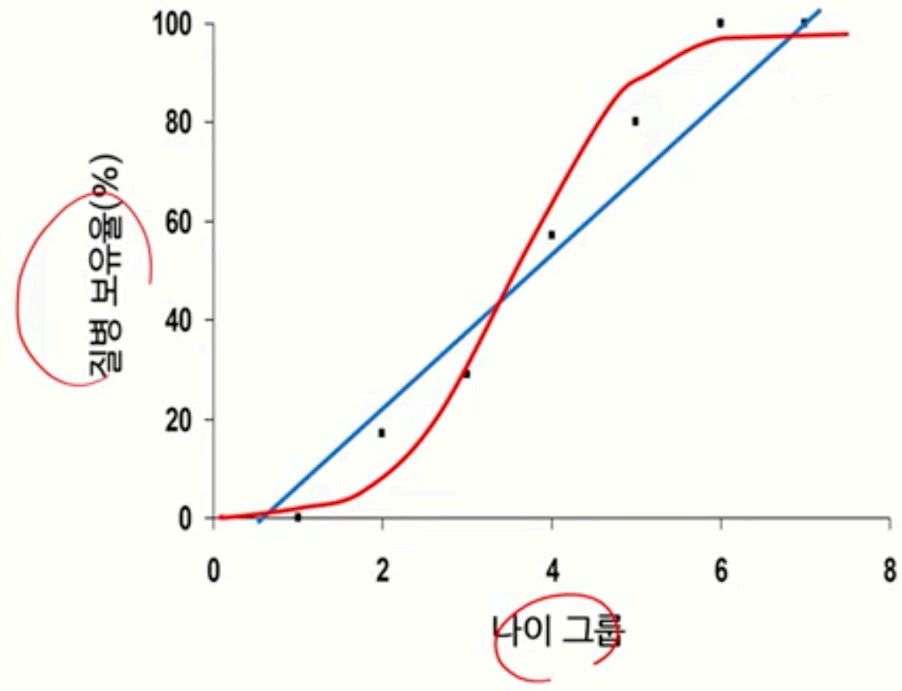
# ◆ Logistic Regression




**굉장히 특징 있는 함수**

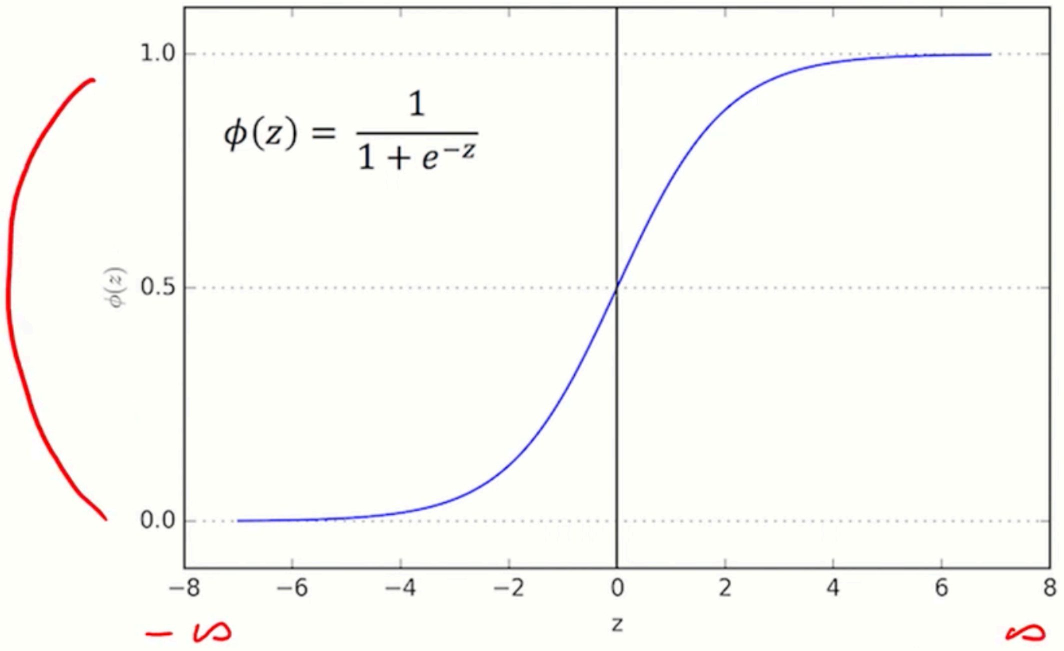
$$f(x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

로지스틱 (Logistic)함수
시그모이드 (Sigmoid)함수



# ◆ Logistic Regression I

- 로지스틱 함수




- Logistic function, Sigmoid function, Squashing function (Large input → Small output)
- 아웃풋 범위: 0~1
- 인풋값에 대해 단조증가 (혹은 단조감소) 함수
- 미분결과를 아웃풋의 함수로 표현 가능 (Gradient learning method에 유용하게 사용)

$$\frac{d\emptyset(z)}{dz} = \frac{1}{1+e^{-z}}\left(1 - \frac{1}{1+e^{-z}}\right) = \emptyset(z)(1- \emptyset(z))$$

# ◆ Logistic Regression I

- 로지스틱 함수

$$E(y) = \pi(X = x) = P(Y = 1|X = x) = 1 - P(Y = 0|X = x)$$

단순로지스틱 회귀모델: 입력변수 X가 1개인 로지스틱 회귀모델

$$E(y) = \pi(X = x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

관측치 $x$가 범주 1에 속할 확률

(Probability that an observation $x$ belongs to class 1)

# ◆ Logistic Regression I

- 로지스틱 함수 $\beta_1$ 해석

$$E(y) = \pi(X = x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

$\beta_1$의 해석 → 직관적이지 못함

▪ 승산 (Odds)

▪ 성공 확률을 $p$ 로 정의할 때, 실패 대비 성공 확률 비율

$$Odd = \frac{p}{1-p}$$

$$p = 1 \rightarrow odd = \infty$$
$$p = 0 \rightarrow odd = 0$$

# ◆ Logistic Regression I

- 로지스틱 함수 $\beta_1$ 해석

FIFA WORLD CUP
RUSSIA 2018

**2018 FIFA WORLD CUP**
**June 14th - July 15th, 2018**
**Various Locations throughout Russia**

| ODDS TO WIN: | OPENING ODDS 7/14/2014 | CURRENT ODDS 2/12/2018 | ODDS TO WIN: | OPENING ODDS 7/14/2014 | CURRENT ODDS 2/12/2018 |
|---|---|---|---|---|---|
| 96101 GERMANY | 5/1 | 4/1 | 96117 SWEDEN | 80/1 | 100/1 |
| 96102 BRAZIL | 8/1 | 5/1 | 96118 SERBIA | 100/1 | 150/1 |
| 96103 FRANCE | 10/1 | 11/2 | 96119 SENEGAL | 500/1 | 150/1 |
| 96104 SPAIN | 8/1 | 13/2 | 96120 EGYPT | 500/1 | 200/1 |
| 96105 ARGENTINA | 8/1 | 7/1 | 96121 ICELAND | 1000/1 | 200/1 |
| 96106 BELGIUM | 15/1 | 10/1 | 96122 PERU | 500/1 | 200/1 |
| 96107 ENGLAND | 25/1 | 15/1 | 96123 NIGERIA | 150/1 | 200/1 |
| 96108 PORTUGAL | 30/1 | 20/1 | 96124 JAPAN | 150/1 | 300/1 |
| 96109 URUGUAY | 50/1 | 30/1 | 96125 COSTA RICA | 200/1 | 250/1 |
| 96110 COLOMBIA | 20/1 | 40/1 | 96126 AUSTRALIA | 300/1 | 300/1 |
| 96111 RUSSIA | 20/1 | 40/1 | 96127 MOROCCO | 500/1 | 500/1 |
| 96112 CROATIA | 60/1 | 40/1 | 96128 IRAN | 2000/1 | 500/1 |
| 96113 POLAND | 100/1 | 40/1 | 96129 SOUTH KOREA | 200/1 | 500/1 |
| 96114 MEXICO | 50/1 | 60/1 | 96130 TUNISIA | 500/1 | 1000/1 |
| 96115 DENMARK | 100/1 | 100/1 | 96131 PANAMA | 1000/1 | 1000/1 |
| 96116 SWITZERLAND | 80/1 | 100/1 | 96132 SAUDI ARABIA | 1000/1 | 1000/1 |

*Odds current as of 1/22/18*

# ◆ Logistic Regression I

- 로지스틱 함수 $\beta_1$ 해석

$$\pi(X = x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}} \qquad 0 \le \pi(X = x) \le 1$$

$$Odds = \frac{\overset{1}{\pi(X = x)}}{\underset{0}{1 - \pi(X = x)}}$$

Odds: 범주 0에 속할 확률 대비 범주 1에 속할 확률

(The ratio of the probability of belonging to class 1 to the probability of belonging to class 0)

# ◆ Logistic Regression I

- 로지스틱 함수 $\beta_1$ 해석

$$\pi(X=x)=\frac{1}{1+e^{-(\beta_0+\beta_1 x)}} \qquad 0\le \pi(X=x)\le 1$$

$$Odds=\frac{\pi(X=x)}{1-\pi(X=x)}$$

1

0

Odds: 범주 0에 속할 확률 대비 범주 1에 속할 확률

(The ratio of the probability of belonging to class 1 to the probability of belonging to class 0)

odds

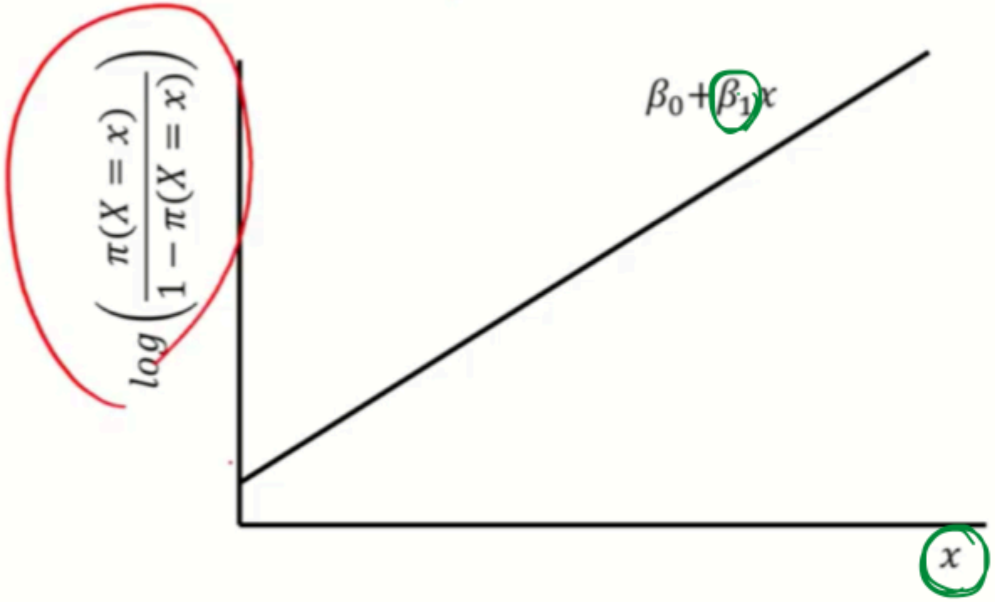
$$log(Odds)=log\left(\frac{\pi(X=x)}{1-\pi(X=x)}\right)=log\left(\frac{\frac{1}{1+e^{-(\beta_0+\beta_1 x)}}}{1-\frac{1}{1+e^{-(\beta_0+\beta_1 x)}}}\right)=\beta_0+\beta_1 x$$

Logit Transform (로짓 변환)

# ◆ Logistic Regression I

- 로지스틱 함수 $\beta_1$ 해석

$$\log\left(\frac{\pi(X=x)}{1-\pi(X=x)}\right) = \beta_0 + \beta_1 x$$




$\beta_1$의 의미: x가 한단위 증가 했을 때 log(odds)의 증가량

# ◆ Logistic Regression

- 회귀 계수의 해석

✓ Linear Regression

- 해당 변수가 1만큼 증가함에 따른 종속변수의 변화량

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 \cdots + \hat{\beta}_d x_d$$

✓ Logistic Regression

- 해당 변수가 1만큼 증가함에 따른 로그 오즈의 변화량

$$log(Odds) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 \cdots + \hat{\beta}_d x_d$$

# ◆ Logistic Regression I

- 로지스틱 함수 $\beta_1$ 해석

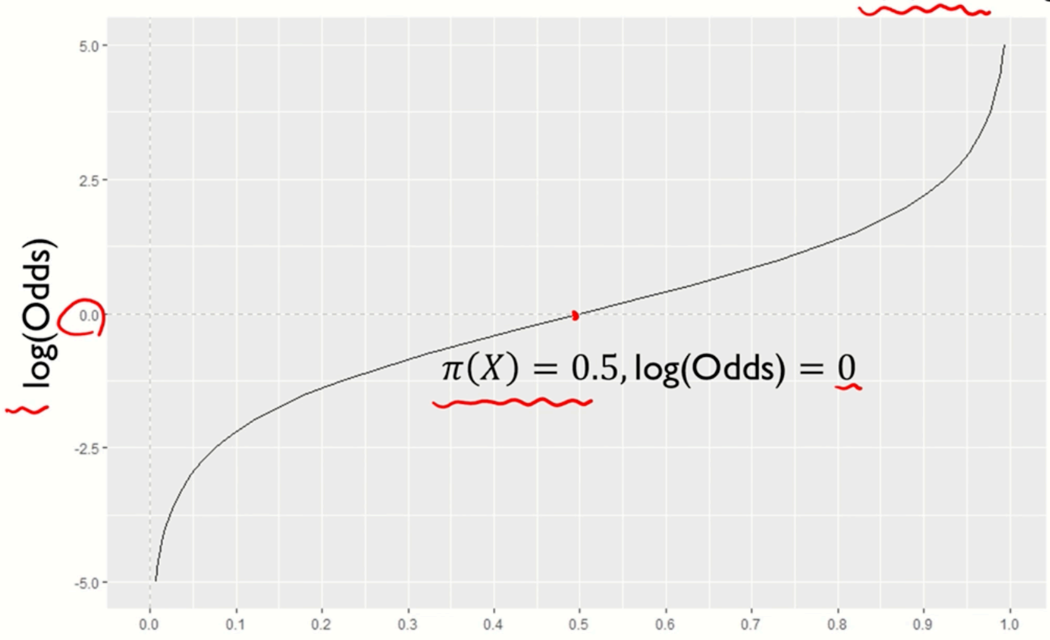
- 성공 확률 $\pi(X)$ 에 따른 log(Odds)의 그래프

$\log\left(\frac{\pi(x)}{1-\pi(x)}\right)$ 0.5 0.5

$\pi(X) \rightarrow 1, \log(\text{Odds}) \rightarrow \infty$

$\pi$

$\log 1 = 0$




$\pi(X) \rightarrow 0, \log(\text{Odds}) \rightarrow -\infty$

$\pi(X)$

## ◆ Logistic Regression

- Logistic Regression – 종속변수 자체가 아닌 **종속변수에 대한 Logit함수**를 **회귀식의 종속변수**로 사용
  - ✓ Odds의 Logit값을 종속변수로 사용한 회귀식을 추정한 뒤, 역산을 통해 구한 성공 확률 값을 기준으로 분류 값을 예측함

$$log(Odds) = log\left(\frac{p}{1-p}\right) = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 \cdots + \hat{\beta}_d x_d$$

양변 exp 취함

$$\frac{p}{1-p} = e^{\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 \cdots + \hat{\beta}_d x_d}$$

P(성공확률)
에 대해 정리

$$p = \frac{1}{1 + e^{-(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 \cdots + \hat{\beta}_d x_d)}}$$

$$= \sigma(\mathrm{x}, \beta)$$

Unit 02 | Logistic Regression

## ◆ Logistic Regression I

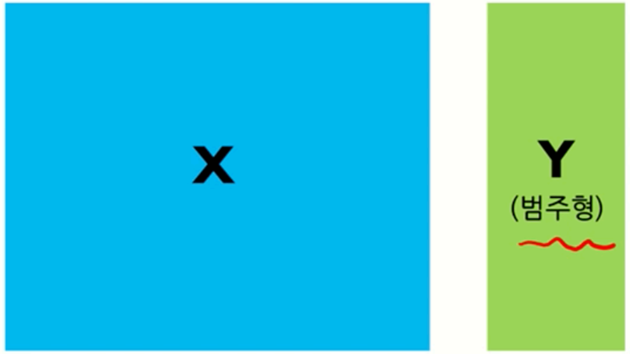
- 로지스틱 회귀모델 배경
- 로지스틱 회귀모델 형태
- Odds

## ◆ Logistic Regression II

- 파라미터 추정
- 로지스틱 회귀모델 결과 및 해석
- 로지스틱 회귀모델 예제

# ◆ Logistic Regression II

- 로지스틱 함수




$$\pi(X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p)}}$$

$$0 \leq \pi(X) \leq 1$$

- X변수를 로지스틱 함수형태 (비선형결합)로 표현
- 관측치가 특정 범주에 속할 확률로 계산
- 확률값이 정한 기준값 보다 크면 범주 1 아니면 범주 2 (이진범주 분류 문제의 경우)

# Logistic Regression II

- 로지스틱 함수

$$\pi(X) = \frac{1}{1 + e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_p X_p)}}$$

$$Odds = \frac{\pi(X)}{1-\pi(X)} = e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_p X_p}$$

1

0

- Logit Transform

$$\log(Odds) = \log\left(\frac{\pi(X)}{1-\pi(X)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$$

## ◆ Logistic Regression II

- 로지스틱 회귀모델 학습
- 수치최적화

- 로지스틱 회귀모델 학습: 최대 우도 추정법 (Maximum Likelihood Estimation)

$$f_i(y_i) = \pi(x_i)^{y_i}(1-\pi(x_i))^{1-y_i}, i = 1,2,\cdots,n$$

$$P(y_i = 1) = \pi_i$$
$$P(y_i = 0) = 1 - \pi_i$$

$$L = \prod_i f_i(y_i) = \prod_i \pi(x_i)^{y_i}(1-\pi(x_i))^{1-y_i}$$

# ◆ Logistic Regression II

- 로지스틱 회귀모델 학습
  - 로지스틱 회귀모델 학습: 최대 우도 추정법 (Maximum Likelihood Estimation)

확률함수

$$f_i(y_i) = \pi(x_i)^{y_i}(1-\pi(x_i))^{1-y_i},\ i = 1,2,\cdots,n$$

$$P(y_i = 1) = \pi_i$$
$$P(y_i = 0) = 1 - \pi_i$$

$$L = \prod_i^n f_i(y_i) = \prod_i \pi(x_i)^{y_i}(1-\pi(x_i))^{1-y_i}$$

log(A B) = log A + log B

$$\ln L = \ln\left[\prod_i \pi(x_i)^{y_i}\left(1-\pi(x_i)\right)^{1-y_i}\right]$$

$$= \ln \prod_i \left[\frac{\pi(x_i)}{1-\pi(x_i)}\right]^{y_i} + \sum_i \ln(1-\pi(x_i))$$

$$= \sum_i y_i \ln\left[\frac{\pi(x_i)}{1-\pi(x_i)}\right] + \sum_i \ln(1-\pi(x_i))$$

$$= \sum_i y_i(\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p) + \sum_i \ln(1 + e^{\beta_0+\beta_1X_1+\cdots+\beta_pX_p})$$

## ◆ Logistic Regression II

- 로지스틱 회귀모델 학습

- 로지스틱 회귀모델 학습: 최대 우도 추정법 (Maximum Likelihood Estimation)

✓ $\ln L = \sum_i y_i(\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p) + \sum_i \ln(1 + e^{\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p})$

- 위 로그-우도함수 (log likelihood function)가 최대가 되는 파라미터 β 결정
- 로그-우도함수 (log likelihood function)는 파라미터 β에 대해 비선형이므로 선형회귀 모델과 같이 명시적인 해가 존재하지 않음 (No closed-form solution exists)
- Iterative reweight least square, Conjugate gradient, Newton's method 등의 수치 최적화 알고리즘을 이용하여 해를 구함

# ◆ Logistic Regression II

- 로지스틱 회귀모델 학습

- Cross entropy: 두 확률분포 ($p(x), q(x)$)의 차이

$$\text{Cross entropy} = -\sum p(x) \log q(x)$$

- Cross entropy: 음의 log likelihood function의 기대값

- Log likelihood function을 최대 = 입력 분포 $p(x)$와 파라미터가 주어졌을 때, 출력 분포 $q(x)$의 확률을 최대

- Cross entropy를 최소 = 입력 분포 $p(x)$와 출력분포 $q(x)$의 차이를 최소

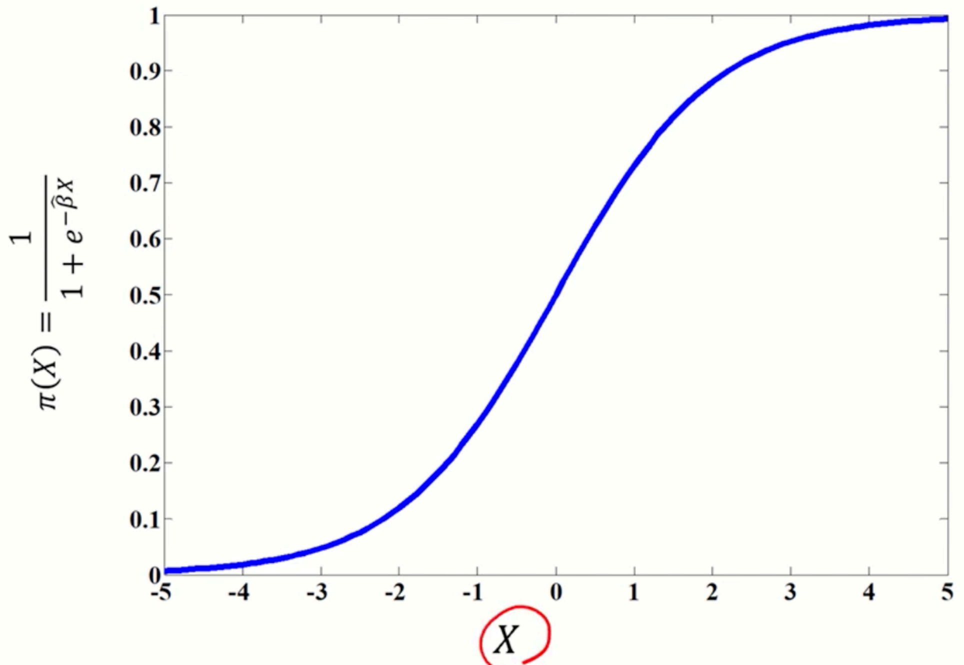
- Log likelihood function을 최대 = cross entropy를 최소

# ◆ Logistic Regression II

- 파라미터 추정 후 최종 모델

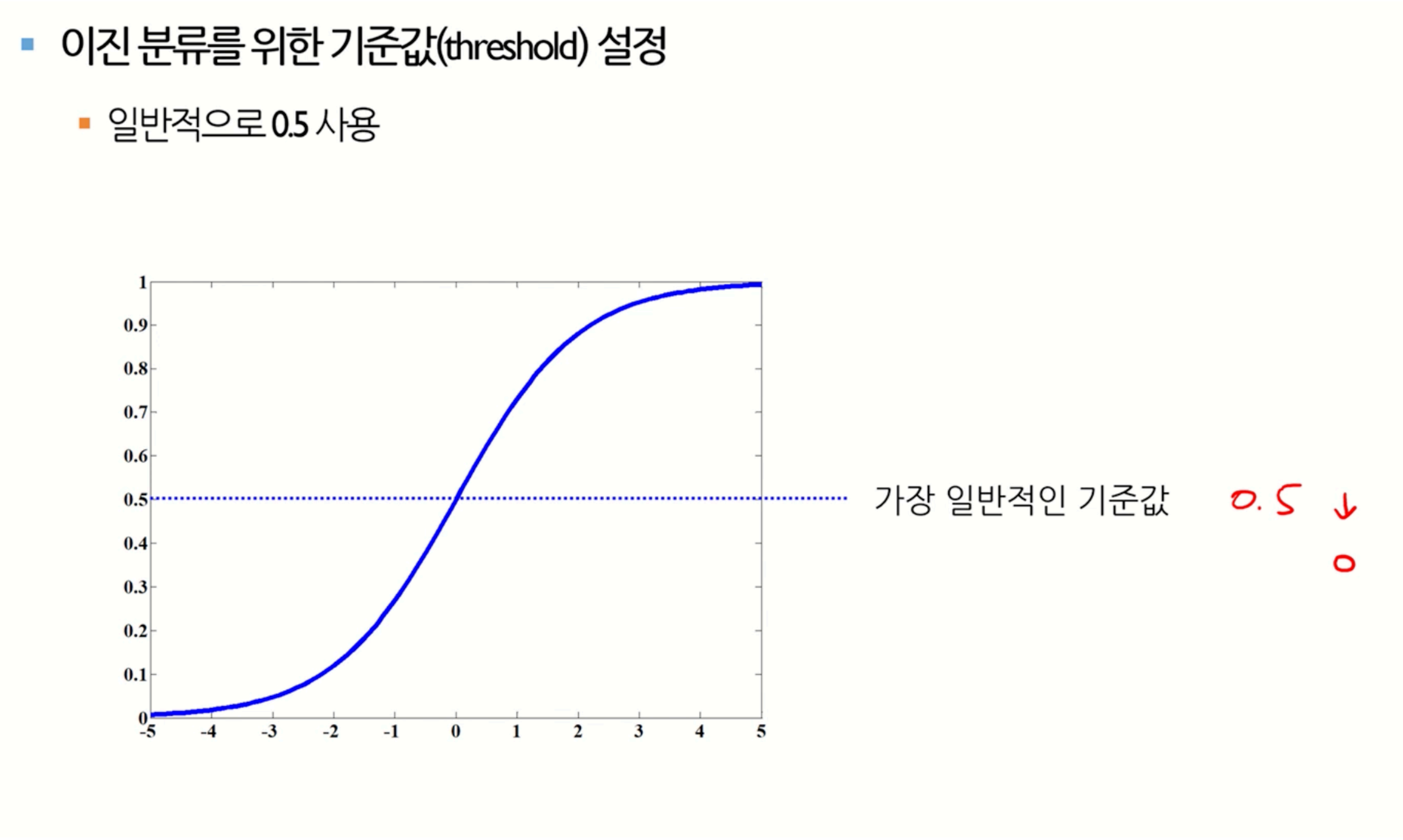
❖ 파라미터가 추정되고 난 이후 최종모델

$$\pi(X) = f(X) = \frac{1}{1 + e^{-(\widehat{\beta_0} + \widehat{\beta_1} X_1 + \cdots + \widehat{\beta_p} X_p)}} = \frac{1}{1 + e^{-\widehat{\beta} X}}$$



# ◆ Logistic Regression II

- 파라미터 추정 후 최종 모델

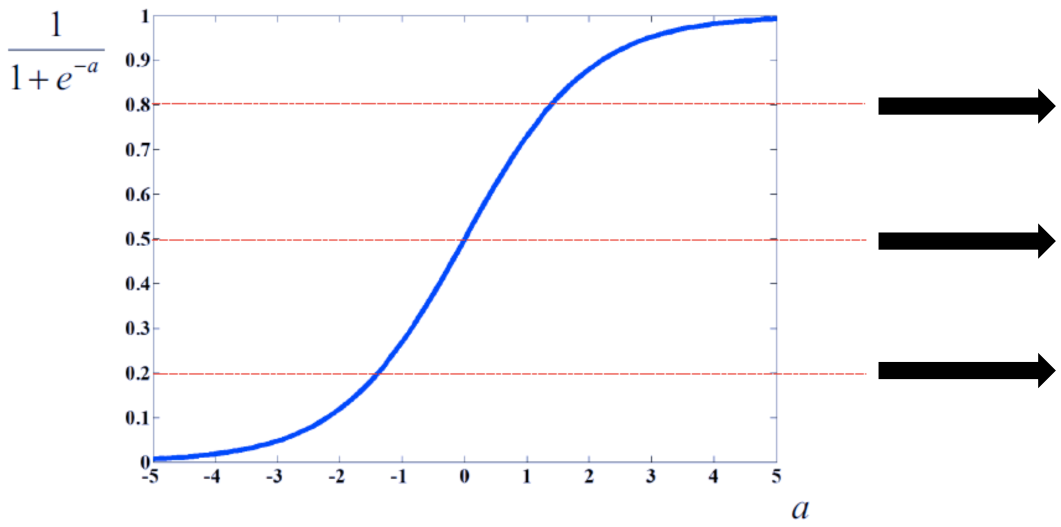


# ◆ Logistic Regression

- Cut-off 설정

✓ 회귀계수가 추정되면 설명변수가 주어졌을 때, 성공확률을 구할 수 있고, 해당 확률이 cut-off 이상이면 1, 아니면 0으로 분류할 수 있음

✓ 사전확률을 고려한 cut-off 또는 검증 데이터의 성능을 최대화하는 cut-off 등을 사용함




1의 비중이 높을때

일반적인 cut-off

1의 비중이 낮을때

# ◆ Logistic Regression II

- 승산 비율: Odds Ratio

- 승산 비율: Odds Ratio

$$\frac{odds(x_1+1, x_2, \cdots, x_n)}{odds(x_1, x_2, \cdots, x_n)} = \frac{e^{\widehat{\beta_0}+\widehat{\beta_1}(X_1+1)+\cdots+\widehat{\beta_p}X_p}}{e^{\widehat{\beta_0}+\widehat{\beta_1}X_1+\cdots+\widehat{\beta_p}X_p}} = e^{\beta_1}$$

- 나머지 입력변수는 모두 고정시킨 상태에서 한 변수를 1단위 증가시켰을 때 변화하는 Odds의 비율
- $x_1$이 1단위 증가하면 성공에 대한 승산 비율이 $e^{\beta_1}$ 만큼 변화함
- 회귀 계수가 양수 → 성공확률 증가 (성공확률 ≥ 1)
- 회귀 계수가 음수 → 성공확률 감소 (0 ≤ 성공확률 < 1 )

# ◆ Logistic Regression II

- 로지스틱 회귀분석 결과 및 해석 (대출 여부를 예측하는 데이터)

$$f(X) = \frac{1}{1 + e^{-(\widehat{\beta_0} + \widehat{\beta_1} X_1 + \cdots + \widehat{\beta_{12}} X_{12})}}$$

| Input variables | Coefficient | Std. Error | p-value | Odds |
|---|---|---|---|---|
| Constant term | -13.20165825 | 2.46772742 | 0.00000009 | * |
| Age | -0.04453737 | 0.09096102 | 0.62439483 | 0.95643985 |
| Experience | 0.05657264 | 0.09005365 | 0.5298661 | 1.05820346 |
| Income | 0.0657607 | 0.00422134 | 0 | 1.06797111 |
| Family | 0.57155931 | 0.10119002 | 0.00000002 | 1.77102649 |
| CCAvg | 0.18724874 | 0.06153848 | 0.00234395 | 1.20592725 |
| Mortgage | 0.00175308 | 0.00080375 | 0.02917421 | 1.00175464 |
| Securities Account | -0.85484785 | 0.41863668 | 0.04115349 | 0.42534789 |
| CD Account | 3.46900773 | 0.44893095 | 0 | 32.10486984 |
| Online | -0.84355801 | 0.22832377 | 0.00022026 | 0.43017724 |
| CreditCard | -0.96406376 | 0.28254223 | 0.00064463 | 0.38134006 |
| EducGrad | 4.58909273 | 0.38708162 | 0 | 98.40509796 |
| EducProf | 4.52272701 | 0.38425466 | 0 | 92.08635712 |

# ◆ Logistic Regression

- Performance evaluation index

| Confusion Matrix | | Predict | |
|---|---|---|---|
| | | Positive | Negative |
| Actual | Positive | a | b |
| | Negative | c | d |

✓ Recall(재현율) = $\frac{a}{a+b}$

✓ Precision(정밀도) = $\frac{a}{a+c}$

✓ F1-measure = $\frac{2*recall*precision}{recall+Precision}$

✓ BCR(균형정확도) = $\sqrt{\frac{a}{a+b}*\frac{d}{c+d}}$

✓ ACC(정분류율)=$\frac{a+d}{a+b+c+d}$

# ◆ Logistic Regression

- Performance evaluation index

| Confusion Matrix | | Predict | |
|---|---|---|---|
| | | Positive | Negative |
| Actual | Positive | TP | FN |
| | Negative | FP | TN |

✓ *Specificity* = $\frac{TN}{TN+FP}$ = $\frac{d}{c+d}$ = Negative로 판단한 것 중 실제 Negative 비율

✓ *Sensitive (Recall)* = $\frac{TP}{TP+FP}$ = $\frac{a}{a+b}$ = 원래 Positive 중에서 Positive로 예측한 비율
= True Positive Rate

✓ *1 - Specificity = False Positive Rate*

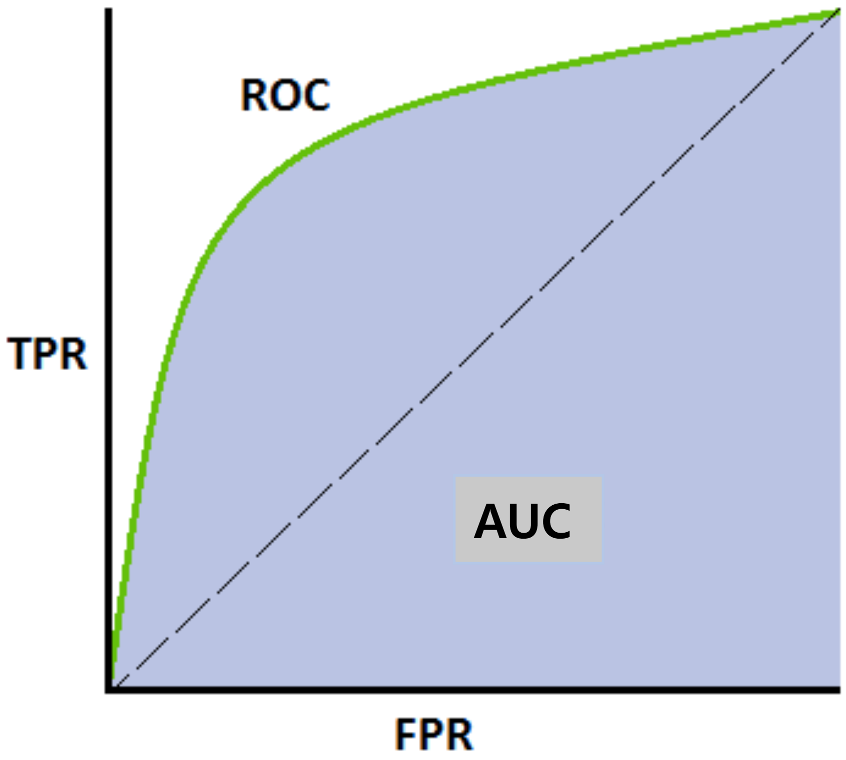
# ◆ Logistic Regression

- ROC Curve

✓ TPR과 FPR은 서로 반비례적인 관계에 있으므로 둘다, **어떤 기준**(언제 1이라고 예측 할 지)을 연속적으로 바꾸면서 측정 해야한다.

✓ 그러면 결국 TPR과 FPR의 여러가지 상황을 고려해서 성능을 판단해야 하는데, 이것을 한눈에 볼 수 있게 한 것이 바로 ROC 커브이다.

✓ AUC : ROC Curve의 밑면적 값, 높을수록 좋다



◆ Logistic Regression II
- 파라미터 추정
- 로지스틱 회귀모델 결과 및 해석
- 로지스틱 회귀모델 예제

## ◆ Logistic Regression II

- 선형 회귀처럼 단순하게 미분을 통해 파라미터를 추정할 수 없었다
- **수치 최적화**를 이용해서 파라미터를 추정했다.
- Log likelihood, Cross entropy …

Contents

Unit 03 | MLE(Maximum Likelihood Estimation)

# ◆ Logistic Regression

- Maximum Likelihood Estimation

✓ 최대우도추정(maximum likelihood estimation)이란 모수(parameter)가 미지의 $\theta$ 인 확률분포에서 뽑은 표본(관측치)들을 바탕으로 확률분포의 모수인 $\theta$ 를 추정하는 기법

✓ 여기에서 우도(likelihood)란 이미 주어진 표본들에 비추어 봤을 때 **모집단의 모수 $\theta$ 에 대한 추정이 그럴듯한 정도**를 가리킴

✓ 따라서, Likelihood $L(\theta|x)$는 $\theta$ 가 전제되었을 때 표본 $x$가 등장할 확률인 $p(x|\theta)$에 비례함

# ◆ Logistic Regression

- Maximum Likelihood Estimation

✓ 최대우도추정(maximum likelihood estimation)이란 모수(parameter)가 미지의 $\theta$ 인 확률분포에서 뽑은 표본(관측치)들을 바탕으로 확률분포의 모수인 $\theta$ 를 추정하는 기법

**무슨 말인지 이해가 잘 가지 않는다.**
**당연하다.**

✓ 여기에서 우도(likelihood)란 이미 주어진 표본들에 비추어 봤을 때 **모집단의 모수 $\theta$ 에 대한 추정이 그럴듯한 정도를** 가리킴

✓ 따라서, Likelihood $L(\theta|x)$는 $\theta$ 가 전제되었을 때 표본 $x$가 등장할 확률인 $p(x|\theta)$에 비례함

# ◆ Logistic Regression

- Maximum Likelihood Estimation

  ✓ 우리는 절대로 모수를 알 수 없다. 따라서 표본(sample)을 통해서 모집단의 특성인 모수를 파악해야한다.

  ✓ 이 때, 우리가 **가정**하는 것은 각각의 표본을 추출할 때의 확률밀도함수 혹은 확률질량함수의 양상을 알고 있다는 것이다. (즉, **수학적으로 각 sample을 뽑을 확률밀도함수에 대해서 어떤 양상을 따르는지 알고 있을 때** 모수를 추정하고자 한다면 최대우도법을 쓸 수 있다.)

  *Q. Bernoulli Random Variable?*

# ◆ Logistic Regression

- Maximum Likelihood Estimation

✓ 어떤 모수 $\theta$ 로 결정되는 확률변수들의 모임 $D_\theta=(X_1, X_2,\cdots, X_n)$ 이 있고 $D_\theta$의 확률밀도함수나 확률질량함수가 $f(x)$이고, 그 확률변수들에서 각각 값 $X_1, X_2,\cdots, X_n$을 얻었을 경우 Likelihood $\boldsymbol{L}(\theta)$는 다음과 같다.

$$\boldsymbol{L(\theta) = f_\theta(x_1, x_2,\cdots, x_n)}$$

Q. 확률 변수에서 각각 값 p(X=x1), p(X=x2)... 를 얻는 게 아닌가요?

✓ 여기에서 Likelihood를 최대로 만드는 $\theta$ 는

$$\boldsymbol{\hat{\theta} = argmax_\theta L(\theta)}$$ 가 된다.

# ◆ Logistic Regression

- Maximum Likelihood Estimation

✓ 이 때 $X_1, X_2, \cdots, X_n$ 이 모두 독립적이고 같은 확률분포를 가지고 있다면 $\boldsymbol{L}$ 은 다음과 같이 표현이 가능하다

$$\boldsymbol{L(\theta)} = \prod_{i}^{n} \boldsymbol{f_{\theta}}(\boldsymbol{x_i})$$

✓ 또한, 로그함수는 단조 증가하므로, $\boldsymbol{L}$ 에 로그를 씌운 값의 최댓값은 원래 값 $\hat{\boldsymbol{\theta}}$와 같고, 이 경우 계산이 비교적 간단해진다.

$$\boldsymbol{L^{*}(\theta)} = \log(\boldsymbol{L(\theta)}) = \sum_{i}^{n} \log \boldsymbol{f_{\theta}}(\boldsymbol{x_i})$$

# ◆ Logistic Regression

- Maximum Likelihood Estimation

✓ 예를 들어 투빅스 11기 12기 중에서 한명을 표본으로 추출하는데 추출된 사람이 남자인지 여자인지를 알려고 한다고 하면, 이 때 표본 확률변수가 갖는 확률분포는 베르누이 분포

$$f(X) = p^X(1-p)^{1-X}$$

✓ 그러면 총 n명에 대해 추출했을 때의 likelihood는 다음과 같이 정해진다.

$$L(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n | p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} \qquad (a)$$

# Unit 03 | MLE(Maximum Likelihood Estimation)

## ◆ Logistic Regression

- Maximum Likelihood Estimation

✓ 가령 10명의 사람을 추출했는데 1번부터 10번 사람까지의 성별이 각각 {남, 여, 남, 남, 여, 여, 남, 남, 여, 남} 이라고 해보자.

✓ 남자라면 $\boldsymbol{X_i = 0}$이라고 하고 여자라면 $\boldsymbol{X_i = 1}$이라고 결정한다고 했을 때, 현 상태에서 $\boldsymbol{x_i(i = 1, 2, \ldots, 10)}$ 은 {0, 1, 0, 0, 1, 1, 0, 0, 1, 0} 이라고 할 수 있다.

그러면 식 (a)는 다음과 같을 것이다.

$$\boldsymbol{L(X_1 = 0, X_2 = 1, \ldots, X_{10} = 0 | p) = \prod_{i=1}^{10} p^{x_i} (1-p)^{1-x_i} \ldots}$$

Unit 03 | MLE(Maximum Likelihood Estimation)

# ◆ Logistic Regression

- Maximum Likelihood Estimation

✓ 그렇다고 해서 우리가 p(표본이 여성일 확률)를 알수 있냐면 그렇진 못하다. 즉, 우리가 정작 알아야 하는 것은 이 p를 어떻게 결정할 것인가이다.

✓ 이 p를 잘 결정하려면 최대한 모집단의 p와 비슷할수록 잘 결정한 것이라고 할 수 있다. 최대우도법에서의 기본 가정은 지금 이러한 상황은 랜덤하게 부여된 것인데,

✓ **이 상황이 나온 것은 이렇게 나올 가능성이 가장 높았기 때문일 것이라는 것이다.**

✓ 따라서, '이 상황'을 설명하는 확률 $f$를 최대화 할 수 있는 모수 $p$를 찾는 것이 최대우도법이 시행하고자 하는 바이다.

## Unit 03 | MLE(Maximum Likelihood Estimation)

# ◆ Logistic Regression

- Maximum Likelihood Estimation

✓ 예시에서 제기된 문제를 계속 풀어보도록 하자. 그런데, 식 (a)에서 함수 $f$를 $p$에 대해 편미분 하려면 쉽지 않다.

✓ 우리는 여기서 로그 함수의 단조증가 성질을 활용하여 $L^* = \log(L)$라는 보조 방정식을 도입하도록 하자.

그러면 $L^*$는 다음과 같다.

$$\boldsymbol{L^* = log(L)} = \sum_{i=1}^{n} \log(\boldsymbol{p}^{\boldsymbol{x_i}}(\mathbf{1}-\boldsymbol{p})^{\mathbf{1}-\boldsymbol{x_i}}) = \sum_{i=1}^{n} \{\boldsymbol{x_i} \log(\boldsymbol{p}) + (\mathbf{1}-\boldsymbol{x_i}) \log(\mathbf{1}-\boldsymbol{p})\}$$

# ◆ Logistic Regression

- Maximum Likelihood Estimation

✓ 이제 보조방정식 $L^*$을 $p$에 대해 편미분 하는 것이 쉬워진다.

✓ 그런 다음 $L^*$ 의 $p$ 에 대한 편미분이 0이 되는 $p$ 를 찾으면 최대우도를 만족하는 모수 $p$ 를 추정할 수 있다.

$$L^* = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{\sum_{i=1}^{n}(1-x_i)}{1-p}$$

$$= \frac{n\bar{X}}{p} - \frac{n(1-\bar{X})}{1-p} = 0$$

따라서

$$\Rightarrow \frac{n\bar{X}}{p} = \frac{n(1-\bar{X})}{1-p}$$

$$\Rightarrow (1-p)\bar{X} = (1-\bar{X})p$$

$$\Rightarrow \bar{X} - p\bar{X} = p - \bar{X}p$$

$$\therefore p = \bar{X}$$

# ◆ Logistic Regression

- Maximum Likelihood Estimation

  ✓ 따라서, $p = \bar{X}$로 모수 $p$를 추정하는 것이 적절하다는 것을 알 수 있다.

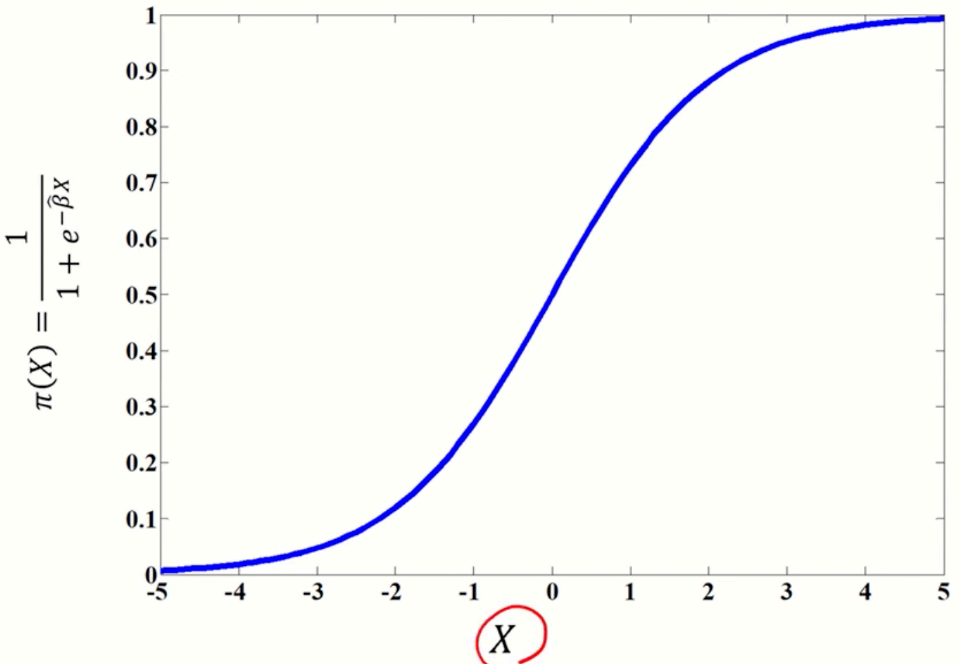
  ✓ 생각해보면 자연스러운 것이 모비율 추정 시 현재 모여있는 사람의 성비를 가지고 모비율을 추정할 수 밖에 없고, 아마 그런 모비율이 있었기 때문에 현재 상태가 만들어 진 것은 아닐까? 라고 추정하는 것은 자연스럽다.

# ◆ Logistic Regression II

• 파라미터 추정 후 최종 모델

❖ 파라미터가 추정되고 난 이후 최종모델

$$\pi(X) = f(X) = \frac{1}{1 + e^{-(\widehat{\beta_0} + \widehat{\beta_1} X_1 + \cdots + \widehat{\beta_p} X_p)}} = \frac{1}{1 + e^{-\widehat{\beta} X}}$$



Contents

# ◆ Optimizer

- Optimizer
  - ✓ Machine learning 문제를 풀다보면 Objective function을 만들고 **그 objective function을 optimize** 해야하는 경우가 매우 빈번하게 발생함.
  - ✓ 간단히 생각해서 **loss function을 minimize**하는 것도 **optimization**이다. <u>그렇다면 그런 optimization은 도대체 어떻게 해야하는 것일까.</u>
  - ✓ 여러가지 방법이 있지만, 이번 시간에서는  간단한 optimization이라는 것에 대한 컨셉을 다루고, 그 중 특수 케이스인 convex optimization에 대해 다룸

Unit 04 | Optimizer

## ◆ Optimizer

- Optimizer
  - ✓ 대부분의 Optimization은 아래와 같은 식으로 표현할 수 있을 것이다.

$$\min f(x)\ s.t.\ g(x) = c$$

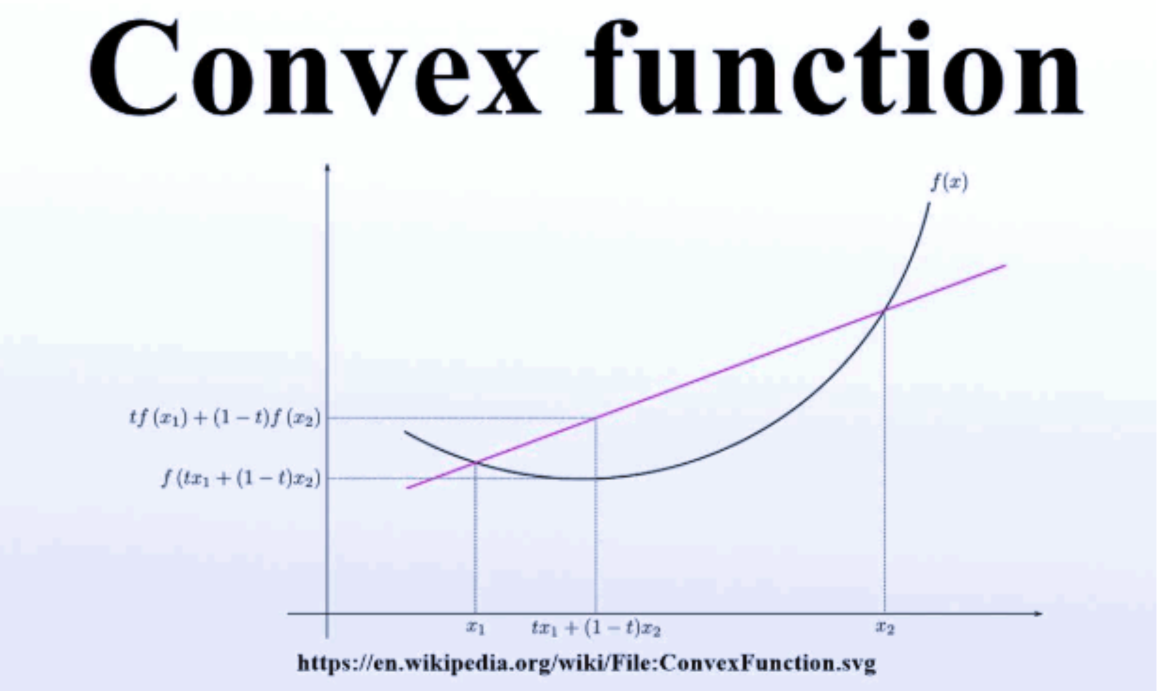
  - ✓ 여기에서 $f(x)$, $g(x)$는 함수이다. 어떤 함수의 optimum point, 즉 그것이 최소이거나 혹은 최대인 지점을 찾는 과정을 optimization이라고 한다고 생각하면 간단
  - ✓ 엄청 간단하게 생각해보면 $f(x)$는 loss function이고, $g(x)$는 일종의 제약조건으로 생각하면 됨

# ◆ Optimizer

- Optimizer
  - ✓ Machine learning 문제를 풀다보면 이렇게 optimization을 해야하는 일이 아주 빈번하게 발생하는데, 안타깝게도 항상 이런 function들의 optimum point를 찾을 수 있는 것은 아님
  - ✓ 가장 간단하게 생각했을 때 이런 point를 찾는 방법은 미분을 하고 그 값이 0이 되는 지점을 찾는 것인데, 안타깝게도 미분 자체가 되지 않는 함수가 존재할 수도 있다
  - ✓ 또한 미분값이 0이라고 해서 반드시 극점인 것은 아니기 때문이다. (saddle point를 생각해보자.)

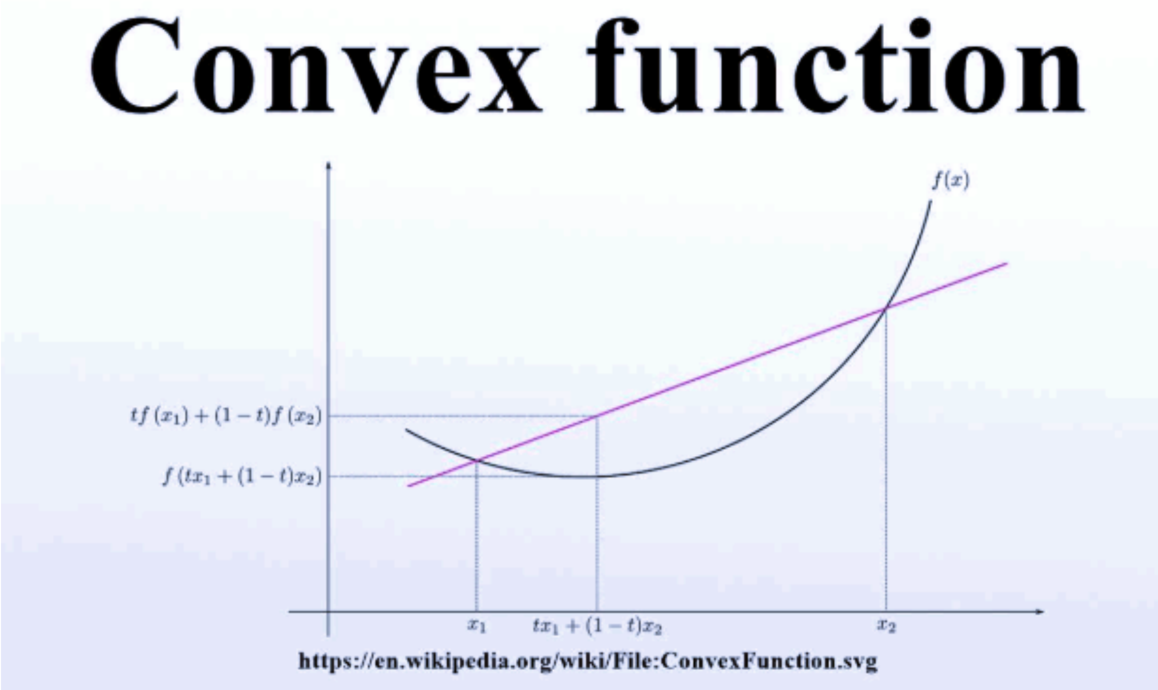
## ◆ Optimizer

- Optimizer
  ✓ 따라서 대부분의 경우에 이런 방법으로 극점을 구하는 것은 불가능하며, 매우매우 특수한 일부 경우에 대해서 완전한 optimum을 찾는 것이 알려져 있다. 그리고 그 경우가 바로 convex optimization이다.



## ◆ Optimizer

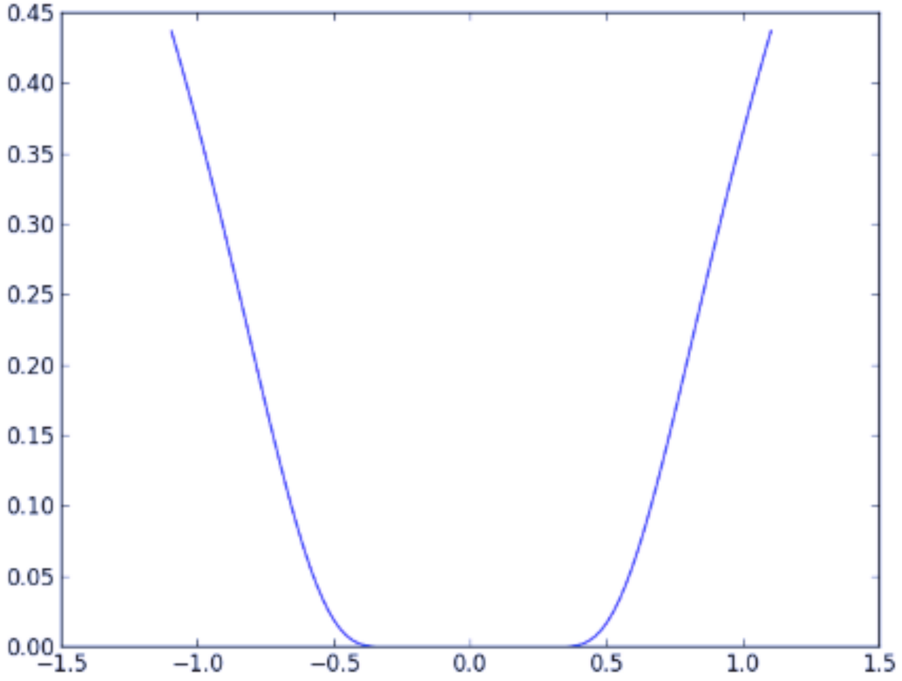
- Optimizer




$$f(\lambda x_1 + (1 - \lambda x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2)\ for\ \forall x_1,\ x_2 \in X,\ \forall \lambda \in [0,1]$$

## ◆ Optimizer

- Optimizer
  - ✓ Convex function이 좋은 이유는 반드시 optimal value가 하나 밖에 존재하지 않는다는 것이다. 여기에서 optimal point가 하나라고 얘기하지 않은 이유는 아래와 같은 예가 있기 때문이다.



# ◆ Optimizer

- Optimizer
  - ✓ 이 함수는 optimal value는 unique하게 존재하지만, 그 값을 가지는 point가 unique하지는 않다.
  - ✓ 따라서 unique한 optimal point를 찾기 위해서는 하나의 조건이 더 필요한데, 바로 strictly convex라는 조건이다.

$$\boldsymbol{f(\lambda x_1 + (1 - \lambda x_2) < \lambda f(x_1) + (1 - \lambda) f(x_2) for \; \forall x_1, \, x_2 \; \in X, \forall \lambda \in [0,1]}$$

  - ✓ 이 조건은 위의 식에서 = 이 빠진 형태이다. 즉 ≤ 가 < 으로 바뀌는 것이다.
  - ✓ 이런 strictly convex function에 대해서 optimal point가 unique하게 존재한다는 것을 증명할 수 있으며, 증명과정은 크게 어렵지 않으니 링크 등을 참고하면 될 것 같다.
  - ✓ 아무튼 strictly convex function은 minimum point가 unique하게 존재하기 때문에, 이런 convex function에 대해서 우리는 어떤 optimization algorithm을 design할 수 있다.

# ◆ Optimizer

- Optimizer

  ✓ 이런 convex optimization의 subset으로 linear programming, quadratic programming, semidefinite programming 등이 존재한다.

  ✓ 여기서는 그런 특수한 경우는 다루지 않고, 일반적인 convex optimization에서 사용할 수 있는 알고리즘들을 다룸.

  ✓ Gradient descent method, Newton method and Lagrange multiplier 등이 있다.

## ◆ Optimizer

- Gradient Descent Algorithm

  ✓ 최적해에 가까워지는 방향으로 조금씩 움직이면서 최적의 해를 찾는 최적화 알고리즘

  ✓ 산 정상 위에 우리가 서있다고 가정

  ✓ 우리가 알 수 있는 정보는 내 위치와 내 주변 위치들의 높이 차이밖에 없다고 가정

  ✓ 만약 내가 산의 가장 낮은 위치로 내려가야하는 상황이라면 어떻게 내려가면 낮은 위치에 도달할 수 있을까?

  ✓ 가장 간단한 방법은 가장 기울기가 가파른 방향을 골라서 내려가는 것

  ✓ 그러다보면 언젠가는 기울기가 0이 되는 지점에 도달하게 될 것이고, 그 지점이 주변에서는 가장 낮은 지점이 될 것이다.

# ◆ Optimizer

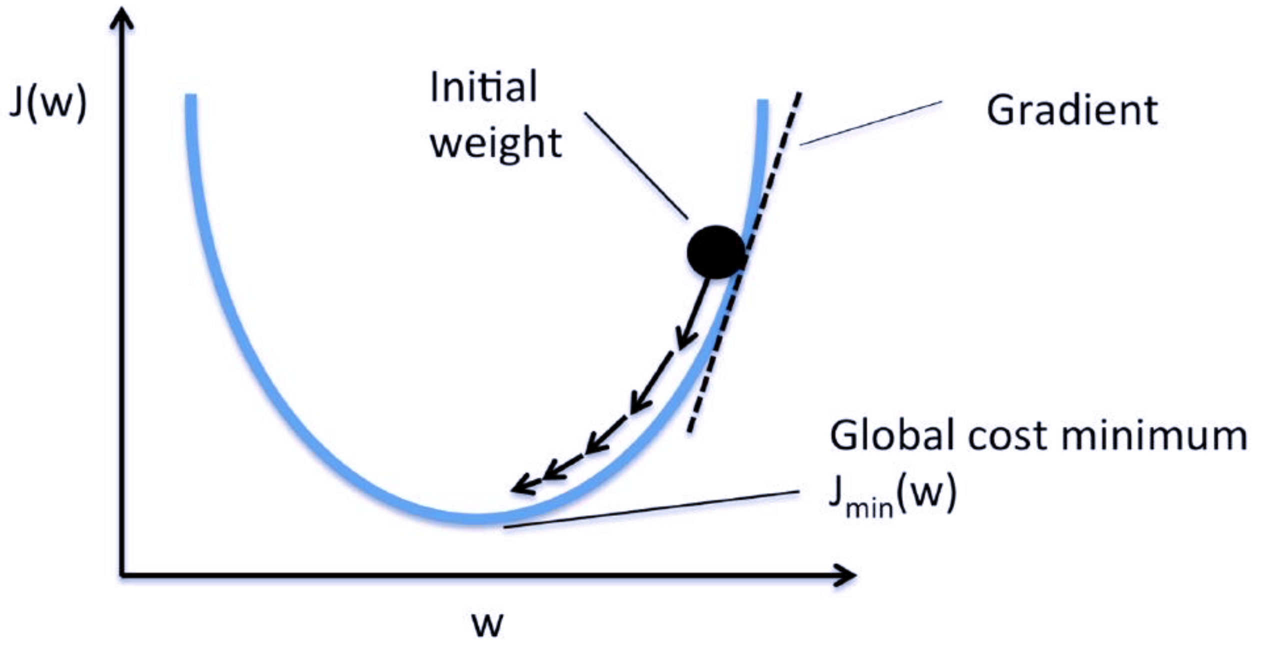
- Gradient Descent Algorithm

  ✓ 그러다보면 언젠가는 기울기가 0이 되는 지점에 도달하게 될 것이고, 그 지점이 주변에서는 가장 낮은 지점이 될 것이다.

  ✓ 만약 산의 높이가 convex function이라면, 즉 가장 낮은 지점이 unique하다면, 그렇게 도달한 지점이 우리가 원했던 가장 optimal한 지점이라는 것을 알 수 있다.

# ◆ Optimizer

- Gradient Descent Algorithm
  - ✓ 파란선 : 가중치에 대한 Loss function
  - ✓ 검은점 : 현재 해의 위치
  - ✓ 화살표 : Loss funstion을 최적화하기 위해 가중치가 이동해야 하는 방향



Unit 04 | Optimizer

# ◆ Optimizer

- Gradient Descent Algorithm – 링크

$$w \leftarrow w - \alpha \frac{\partial}{\partial w} J(w)$$

## ✓ $\frac{\partial}{\partial w} J(w)$: gradient

- 가중치에 대한 목적함수의 미분값
- 미분 값의 부호는 최적해로 가기 위한 방향과 반대 값을 가짐

## ✓ $\alpha$ : Learning Rate

- 움직이는 정도를 조정하는 계수로 클수록 빠르게 학습하지만 최적해로 수렴하지 못할 수도 있음

Contents

# Unit 05 | Tutorial & Assignment

## ◆ Assignment1

- 실습코드 1에서 Q1, Q2 채우기

## ◆ Assignment2

**<Money Ball>** https://www.kaggle.com/wduckett/moneyball-mlb-stats-19622012

1. 어떤 지표가 우리 팀을 Play-off로 이끄는지 학습해보기
   3-1. Moneyball Kernel - Using data from The History…
   3-2. EDA & Preprocessing
   3-3. Logistic Regression

Unit 05 | Tutorial & Assignment

## ◆ Tutorial

• Logistic Regression 맛보기 with Python(Logistic_Regression_01.ipynb)

## ◆ Assignment3

1. 튜토리얼 파일(Logistic_Regression_01.ipynb)을 수정하세요.

   * confusion matrix와 ROC curve를 활용하여 cut-off만 조정
   * 새로운 결과값 도출 및 의미해석

2. MLE를 Gradient Descent의 Loss function으로 사용하여 최적화하기

   * Logistic Regression의 회귀계수를 추정하는 MLE와 Optimizer 중에 하나인 GD를 배웠습니다.
   * MLE를 최적화 해보세요.(MLE에 마이너스를 곱하면 Convex 함수가 됩니다)

## References

## ◆ Reference


- 투빅스 7기 최희정님 자료, 투빅스 7기 전종섭님 자료
- 투빅스 10기 박성진님 자료
- 모두의 딥러닝 머신러닝, 로지스틱회귀
- 고려대학교 김성범 교수님의 머신러닝 로지스틱 회귀 강의 1,2
- Ratsgo's blog
- 조대협님 blog
- Quora
- 공돌이의 수학정리노트
- Lee, I. Optimization. Retrieved from http://issactoast.com
- SanghyukChun's Blog – Convex Optimization
- Kaggle – 타이타닉

<이미지>
http://blog.bidmotion.com/2016/06/23/good-morning-have-you-used-machine-learning/arthur-samuel-quote-machine-learning/

# Q & A