The Apprentice Project

Lec05: Linear Regression (Part 2)

충북대학교

문성태 (지능로봇공학과)

stmoon@cbnu.ac.kr

04

Linear Regression with MLE

(Part I: Probability)

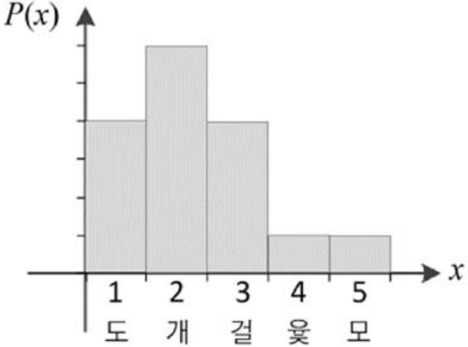
Probability – Random variable

- 확률변수 (random variable)
  - Ex) 윷

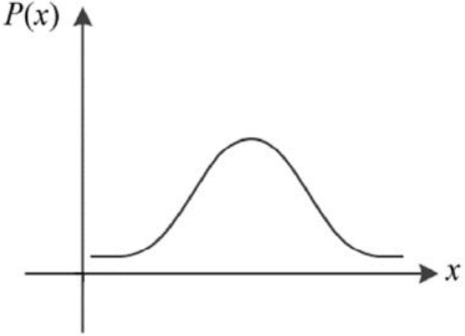
  - 다섯 가지 경우 중 한 값을 갖는 확률변수 x
  - x의 정의역은 {도, 개, 걸, 윷, 모}

Probability – Random variable

- 확률 분포



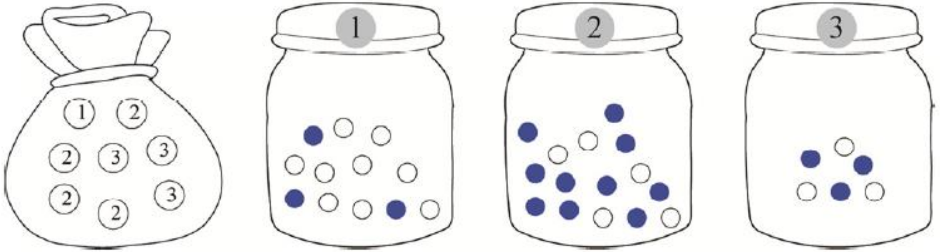
(a) 이산인 경우의 확률질량함수



(b) 연속인 경우의 확률밀도함수

Probability – Example

- 주머니에서 번호를 뽑은 다음, 번호에 따라 해당 병에서 공을 뽑고 색을 관찰함
- 번호를 y, 공의 색을 x라는 확률변수로 표현하면 정의역은 y∈{①,②,③}, x∈{파랑, 하양}



Probability – 곱의 규칙

- 곱의 규칙
  - 두 사건 A, B에 대하여 사건 A가 일어나는 경우의 수가 m이고, 그 각각에 대하여 사건 B가 일어나는 경우의 수가 n이면 두 사건 A, B가 연이어 일어나는 경우의 수는 m*n 이다
  - ①번 카드를 뽑을 확률은 $P(y=①)=P(①)=1/8$
  - 카드는 ①번, 공은 하얀색일 확률은 $P(y=①,x=\text{하얀색}) = P(①,\text{하얀색})$ ← 결합확률

$$\text{곱 규칙: } P(y,x) = P(x|y)P(y)$$

Probability – 합의 규칙

- 합의 규칙
  - 두 사건 A, B가 동시에 일어나지 않을 때, 사건 A와 사건 B가 일어나는 경우의 수가 각각 m, n이면 사건 A 또는 사건 B가 일어나는 경우의 수는 m+n 이다

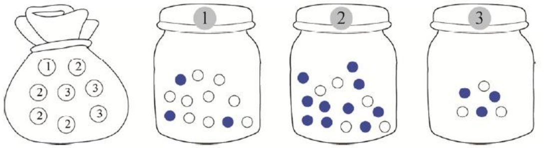
$$\text{합 규칙: } P(x) = \sum_y P(y, x) = \sum_y P(x|y)P(y)$$

Bayes' theorem

$$P(y,x) = P(x|y)P(y) = P(x,y) = P(y|x)P(x)$$

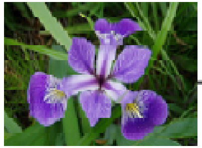
$$P(y|x) = \frac{P(x|y)\ P(y)}{P(x)}$$

Example - Bayes' theorem #1



“하얀 공이 나왔다는 사실만 알고 어느 병에서 나왔는지 모르는데, 어느 병인지 추정하라.”

Example - Bayes' theorem #2



$\xrightarrow{\text{특징추출}}$ $\mathbf{x} = (7.0, 3.2, 4.7, 1.4)^{\mathrm{T}}$ $\xrightarrow{\text{사후확률 추정}}$

$$P(\text{setosa}|\mathbf{x}) = 0.18$$
$$P(\text{versicolor}|\mathbf{x}) = 0.72$$
$$P(\text{virginica}|\mathbf{x}) = 0.10$$

$\xrightarrow{\text{argmax}}$ versicolor

Bayes' theorem

X: 관측치 (Data) : D
θ: 모수 (Population Parameter) : H

$$P(\theta \mid X) = \frac{P(X|\theta)P(\theta)}{P(X)}$$

Likelihood

- Probability
    - 주어진 확률분포가 있을 때, 관측값 혹은 관측 구간이 분포 안에서 얼마의 확률로 존재하는 지를 나타내는 값
- Likelihood
    - 어떤 값이 관측되었을 때, 해당 관측값이 어떤 확률분포로부터 나왔는지에 대한 확률
    - "확률"의 개념과는 반대로 고정되는 요소가 확률분포가 아닌 관측값 D

Maximum Likelihood Estimation (MLE)

- 각 관측값 X에 대한 총 가능도(즉, 모든 가능도의 곱)가 최대가 되게 하는 확률분포를 찾는 것

Maximum Likelihood Estimation (MLE)

- 모든 관측 데이터가 독립적이라고 가정

- 곱셈의 경우 데이터의 개수가 많아질 수록 값이 작아지고, 미분이 어려워 log를 취한 log likelihood 활용
- 미분을 통해 최소값을 구하기 위해 “-” 활용

05

Linear Regression with MLE

(Part II: MLE)

Maximum Likelihood Estimation

- 잔차 항이 정규분포를 따른다고 가정 ➔ $\epsilon \sim N(0, \sigma^2)$

$$\hat{y}_i = w_1 x_{1,i} + w_2 x_{2,i} + \ldots + w_k x_{k,i} + b$$

Maximum Likelihood Estimation

$$\max_{w,b} \log(L(w,b)) = \max_{w,b}\left(-\sum_i \left(y_i - \hat{y}_i\right)^2\right)$$

06

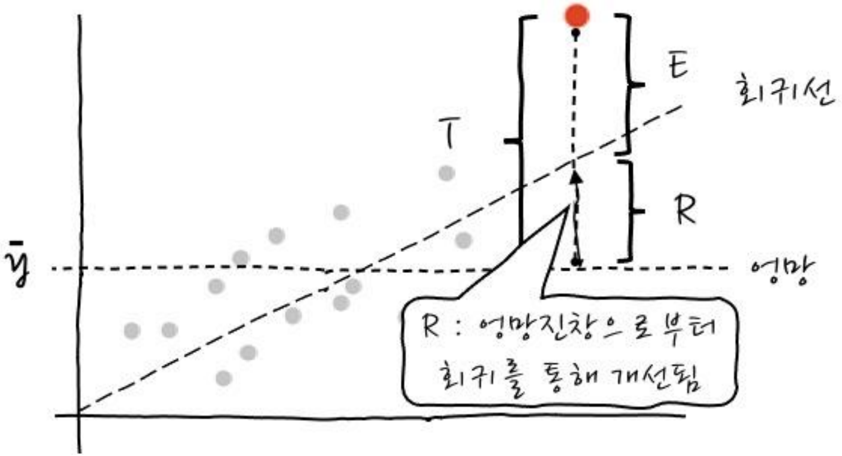
Performance Evaluation for Linear Regression

Performance Evaluation

- 전체 데이터를 학습용과 평가용으로 나눈 다음, 학습용 데이터로 회기평면 추정
- 평가용 데이터에 적용하여 성능 평가
- 평가 방법
  - MSE (Mean Square Error)
  - R^2 (결정 계수)

$$MSE = \frac{1}{n}\sum_{i}^{n}(y_i - \hat{y}_i)^2$$

Coefficient of Determination (R2, 결정계수)

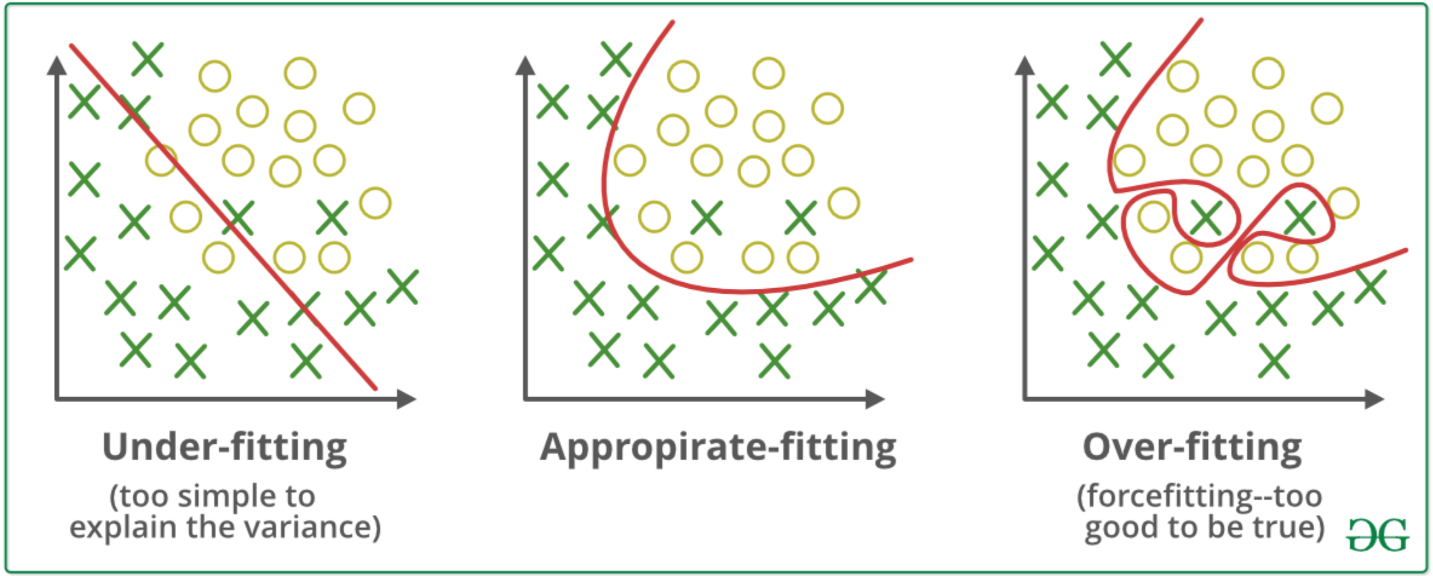


07

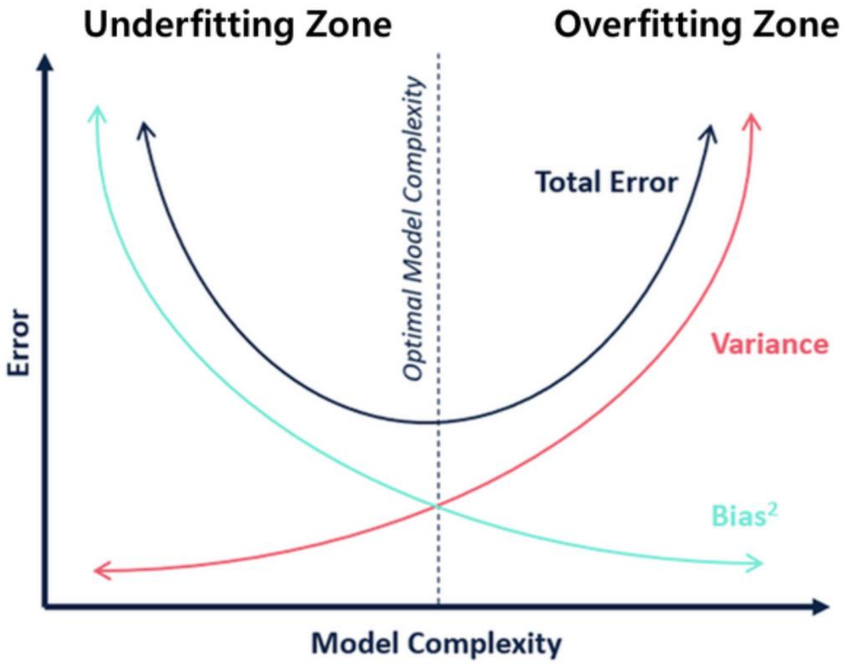
Regularization

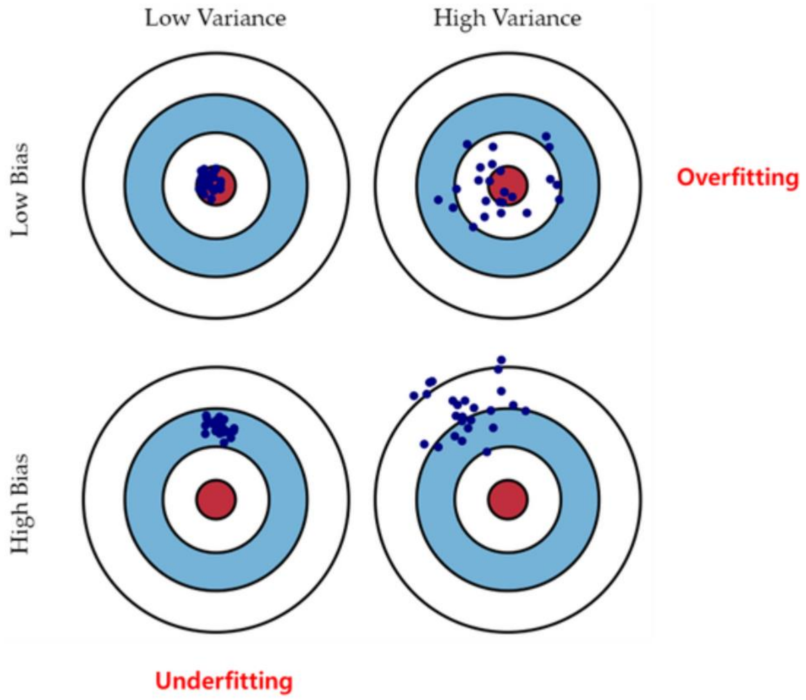
Overfitting

- 학습데이터에 너무 최적화되어 W 값이 잡히고, 이후 학습 데이터가 아닌 새로운 데이터에는 올바른 값을 내보내지 못하는 현상



Overfitting vs Underfitting





https://www.geeksforgeeks.org/underfitting-and-overfitting-in-machine-learning/

Solution for Overfitting

- 학습 데이터 양을 늘리기
- Batch Normalization
- 모델의 복잡도 줄이기
- Drop-Out
- Regularization

Regularization

- 과대 적합을 완화하기 위한 대표적인 방법
- L1 Regularization (ex: Lasso)
- L2 Regularization (ex: Ridge)

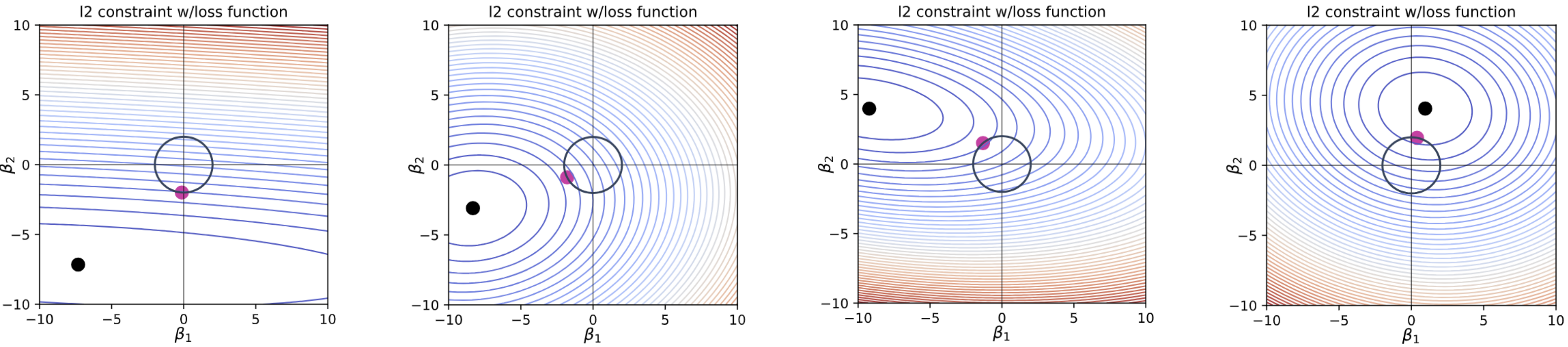
Norm

- N-Norm

- L1/L2 Norm

Ridge Regression

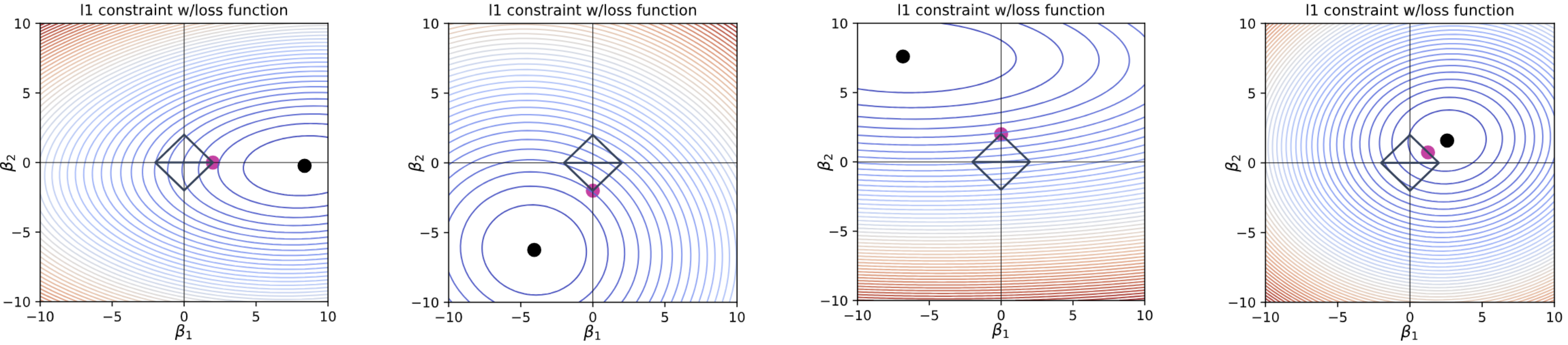
- 계수의 제곱한 값을 기준으로 규제



https://explained.ai/regularization/constraints.html#sec:2.2

Lasso Regression

- 계수의 절대값을 기준으로 규제



특정 feature를 선택하고자 하는 경향이 있음

https://explained.ai/regularization/L1vsL2.html

Lasso vs Ridge

- Lasso (L1)
  - ____________________
  - ____________________
  - ____________________
- Ridge (L2)
  - ____________________
  - ____________________
  - ____________________