



Foundation of Machine Learning 4주차



정재현, 우지수 / 2023.02.10



Computational Data Science LAB



CONTENTS

1. Regression Loss Functions
 2. Classification Loss Functions
 3. Convex Optimization
 4. The Lagrangian
 5. Extra Study
- 
- 

01 | Regression Loss Functions

General Regression Loss Function

✓ 일반적인 손실함수는 아래와 같은 형태

$$(\hat{y}, y) \rightarrow \ell(\hat{y}, y)$$

01 | Regression Loss Functions

Distance Based Regression Loss

✓ Loss $\ell(\hat{y}, y)$ 는 다음과 같은 경우를 따를 때 거리 기반이라고 정의

- Loss가 잔차에 의존할 때

$$\ell(\hat{y}, y) = \psi(y - \hat{y})$$

- Loss는 잔차가 0일 때 0임.

$$\psi(0) = 0$$

* 잔차 $r = y - \hat{y}$ (예측하고자 한 것과 예측한 것의 차이, 정답을 얻기 위해 예측에 추가)

✓ Translation invariant : 함수의 입력 값이 바뀌어도 출력은 그대로 유지되어 바뀌지 않는다는 의미

$$\ell(\hat{y} + a, y + a) = \ell(\hat{y}, y)$$

01 | Regression Loss Functions

Absolute Loss and Square Loss

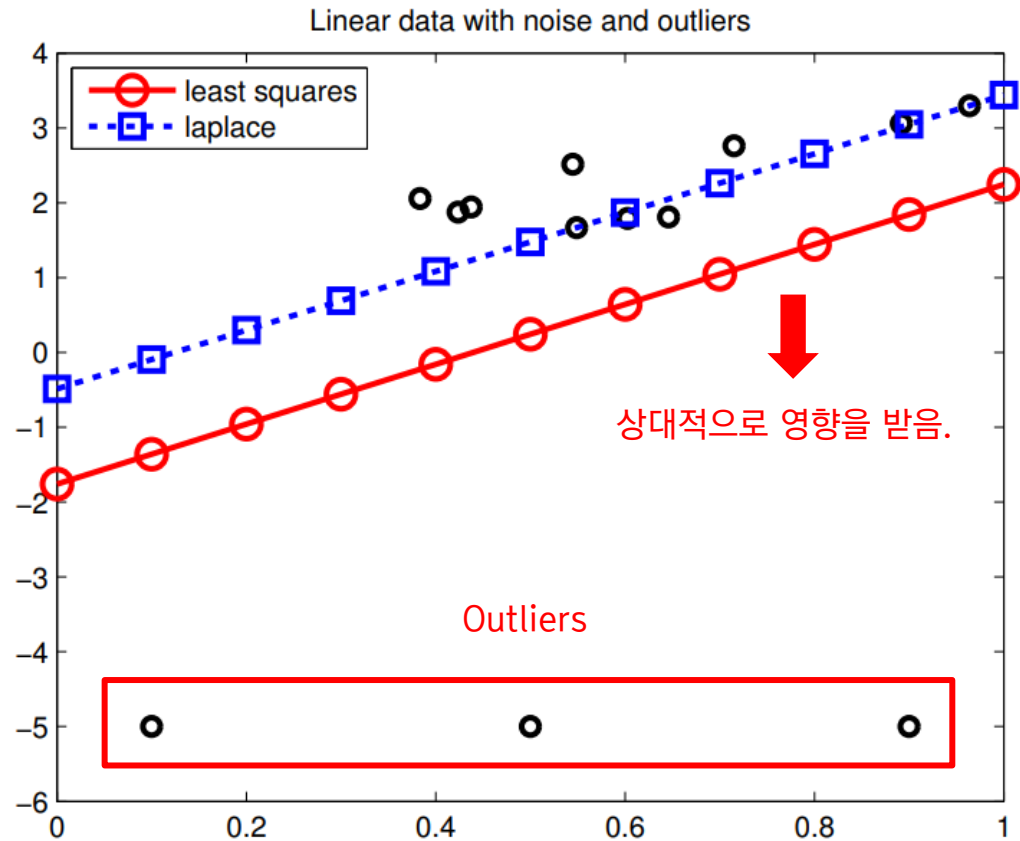
Absolute(Laplace) Loss Square Loss

y	\hat{y}	$ r = y - \hat{y} $	$r^2 = (y - \hat{y})^2$
1	0	1	1
5	0	5	25
10	0	10	100
50	0	50	2500

- ✓ Outliers는 일반적으로 잔차가 큼.
- ✓ Square Loss는 Absolute Loss보다 Outlier의 영향을 많이 받음.

01 | Regression Loss Functions

Loss Function Robustness



Absolute(Laplace) Loss

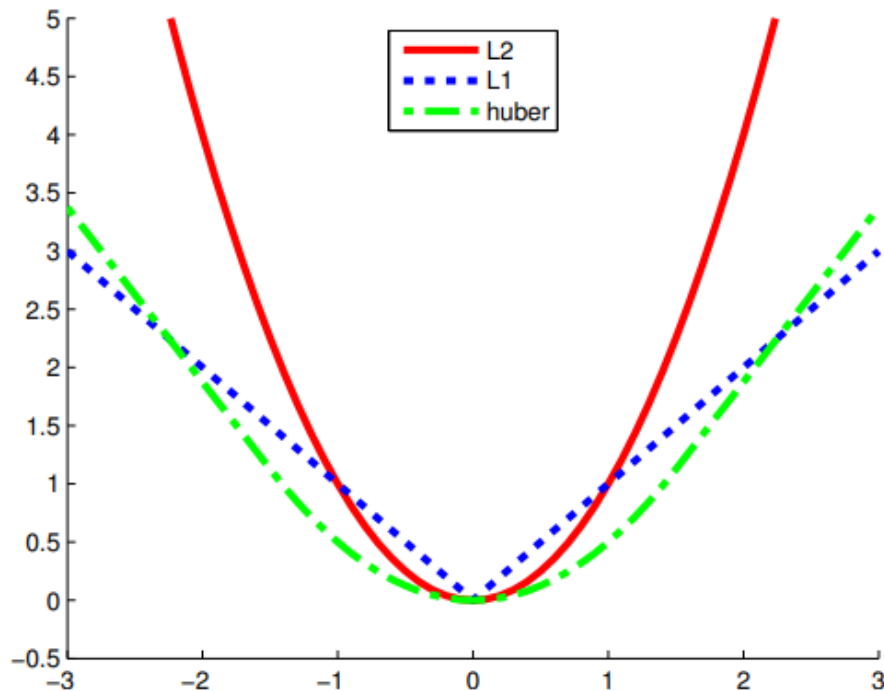
Square Loss

y	\hat{y}	$ r = y - \hat{y} $	$r^2 = (y - \hat{y})^2$
1	0	1	1
5	0	5	25
10	0	10	100
50	0	50	2500

- ✓ Robustness : 학습 알고리즘이 Outlier에 의해 얼마나 영향을 받는지를 나타냄.

01 | Regression Loss Functions

L1, L2, Huber Loss



- ✓ L1 Loss : 미분이 불가능한 지점이 있지만 L2에 비해 상대적으로 Outlier에 대한 영향은 적음.
- ✓ L2 Loss : 모든 지점에서 미분이 가능하지만, Outlier에 대한 영향이 큼.



- ✓ Huber Loss : 모든 지점에서 미분이 가능하며 이상치를 적당히 반영함.

02 | Classification Loss Functions

General Classification Loss Function : The Score Function

✓ Classification Loss Functions

$$\ell(f(x), y) = 1(\underline{f(x)} \neq y)$$

Input x 에 대한 score

* Score size : 예측의 신뢰도

02 | Classification Loss Functions

Distance Based Classification Loss : Margin

✓ Margin

= 예측이 얼마나 정확한지를 보여주는 척도 → 마진을 최대화 시키는 것이 목표

✓ 대부분의 Classification Losses는 마진에 의존

$$y\hat{y} = y\underline{f(x)}$$

Input x 에 대한 score

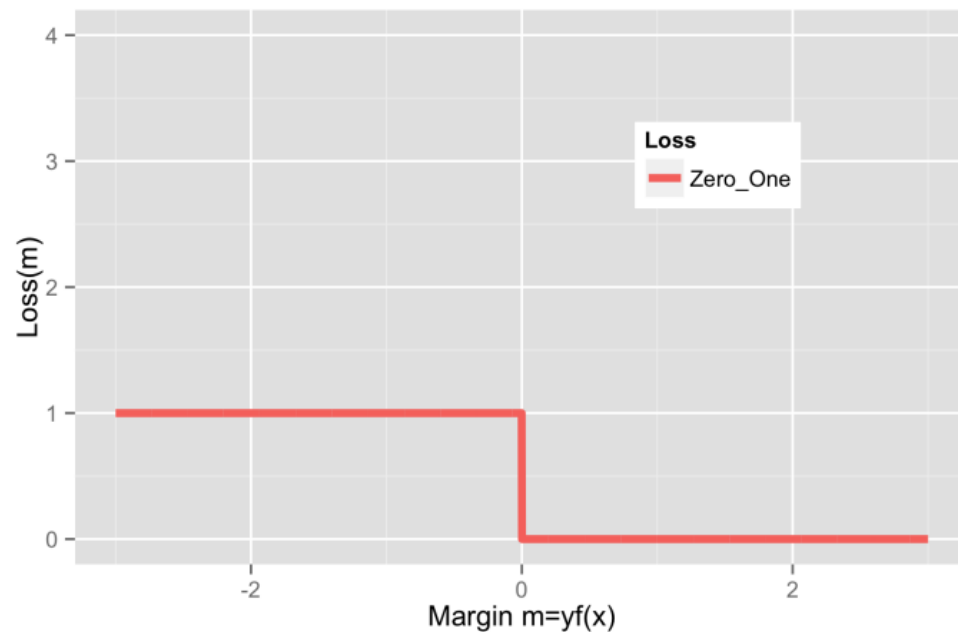
양수 : 예측이 정확할 때

음수 : 예측이 잘못됐을 때

02 | Classification Loss Functions

Zero – One Loss

$$\ell_{0-1} = 1 (m \leq 0)$$

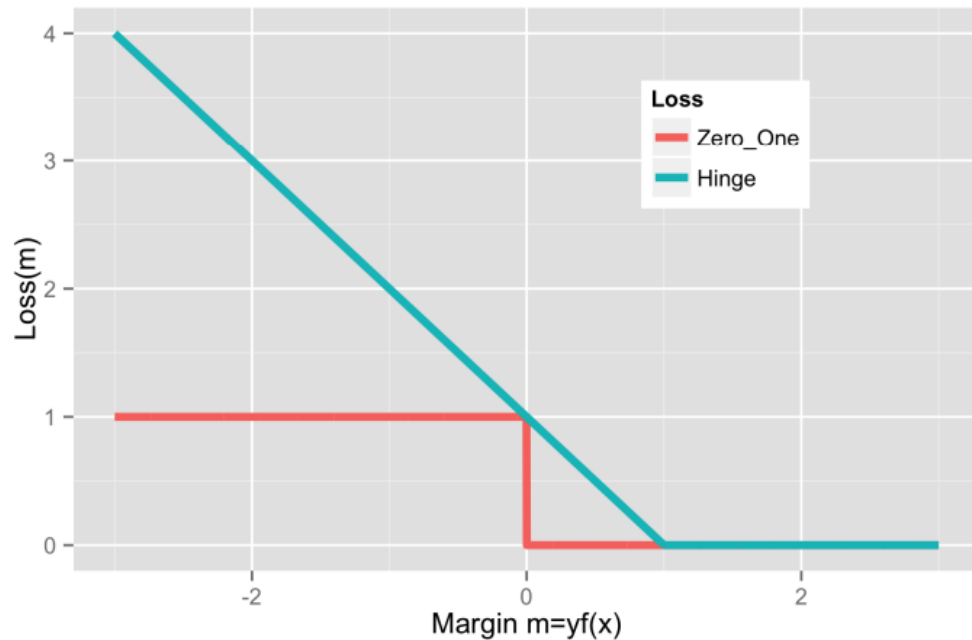


- ✓ Margin ≥ 0 ➡ Loss = 0 (Correct Classification)
- ✓ Margin < 0 ➡ Loss = 1

02 | Classification Loss Functions

Hinge Loss

$$\ell_{Hinge} = \max\{1 - m, 0\} = (1 - m)_+$$

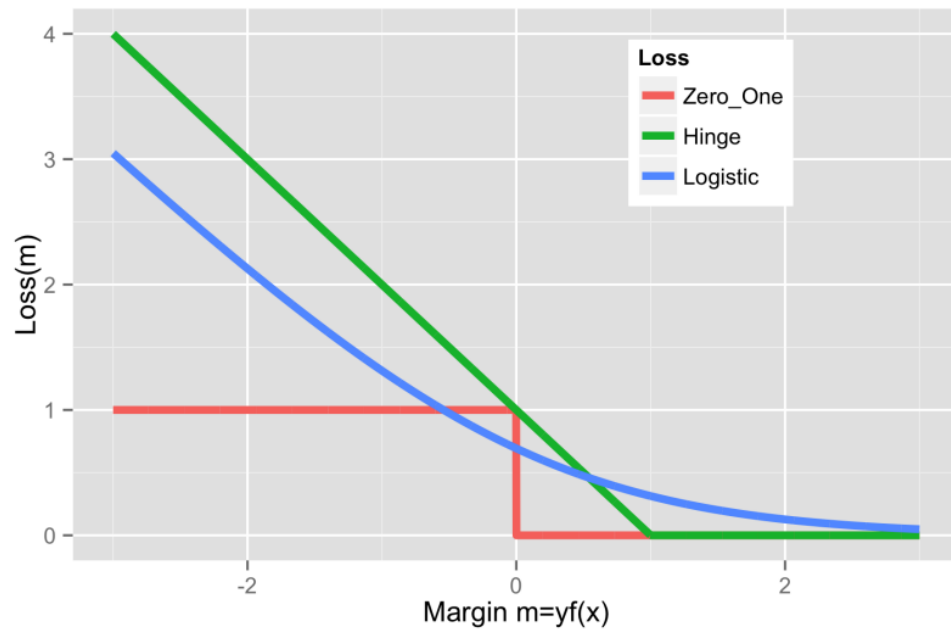


- ✓ Not differentiable (미분 불가능) , Convex (볼록)
- ✓ Margin < 1 ➡ Loss가 선형적으로 증가
- ✓ Margin ≥ 1 ➡ Loss = 0
- ✓ SVM에 활용되는 Loss

02 | Classification Loss Functions

Logistic Loss

$$\ell_{\text{Logistic}} = \log(1 + e^{-m})$$

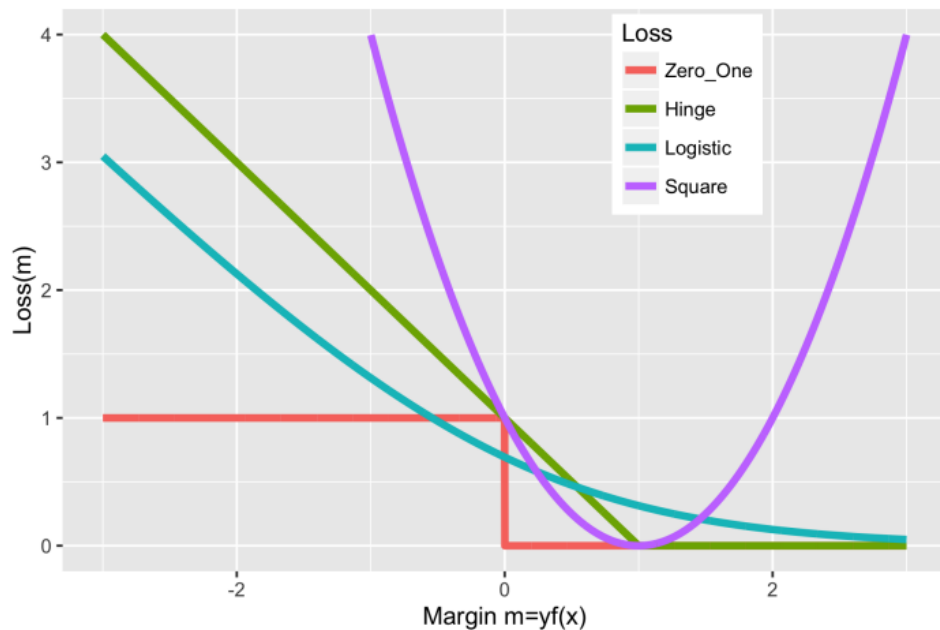


✓ Differentiable (미분가능)

02 | Classification Loss Functions

Square Loss

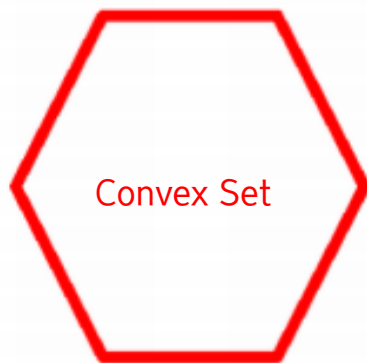
$$\ell(f(x), y) = (f(x) - y)^2 = (1 - f(x)y)^2 = (1 - m)^2$$



✓ Outlier에 대해 영향을 크게 받음.

03 | Convex Optimization

Convex Sets



Convex Set



Not Convex Set

$$\theta x_1 + (1 - \theta)x_2 \in C$$
$$(x_1, x_2 \in C, 0 \leq \theta \leq 1)$$

- ✓ 어떠한 집합이 주어졌을 때, 이 집합의 원소인 x_1, x_2 를 잇는 선분이 이 집합에 다시 포함될 때 해당 집합을 Convex Set이라고 정의

03 | Convex Optimization

Convex Function



Convex Function

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$$

- ✓ 함수 f 의 그래프 상의 임의의 두 점 $(x, f(x)), (y, f(y))$
- ✓ 구간 $[x, y]$ 에서 두 점을 잇는 직선보다 구간 안의 함수 f 값이 더 작을 때, 위 함수 f 는 볼록하다고 정의

03 | Convex Optimization

: When there is Constrained Functions

$$\begin{array}{ll} \text{minimize} & f_0(\textcircled{x}) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \dots, m \\ & h_i(x) = 0, \quad i = 1, \dots, p, \end{array}$$

최적화 변수

↑

목적함수

04 | The Lagrangian

Background

$$L(x, \lambda) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x)$$

λ_i 라그랑주 승수 (dual variable)

$$\begin{aligned} \sup_{\lambda \succeq 0} L(x, \lambda) &= \sup_{\lambda \succeq 0} \left(f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) \right) \\ &= \begin{cases} f_0(x) & \text{when } f_i(x) \leq 0 \text{ all } i \\ \infty & \text{otherwise.} \end{cases} \end{aligned}$$

✓ 제약조건이 존재하는 최적화 문제에서는 라그랑주 승수법을 이용해서 최적화 가능 = primal

04 | The Lagrangian

The Primal and the Dual

1. Primal

$$p^* = \inf_x \sup_{\lambda \geq 0} L(x, \lambda)$$

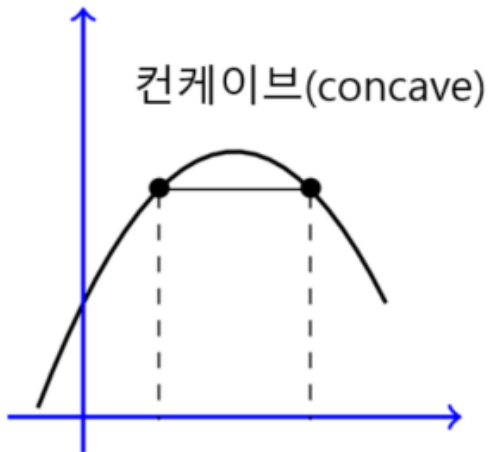
2. Dual (non convex 문제 해결 가능)

$$d^* = \sup_{\lambda \geq 0} \inf_x L(x, \lambda)$$

✓ Inf와 sup의 위치를 바꿈으로써 Lagrangian Dual Problem을 얻을 수 있음.

04 | The Lagrangian

Dual Problem : Search for Best Lower Bound



Dual Problem

$$d^* = \sup_{\lambda \succeq 0} \inf_x L(x, \lambda)$$



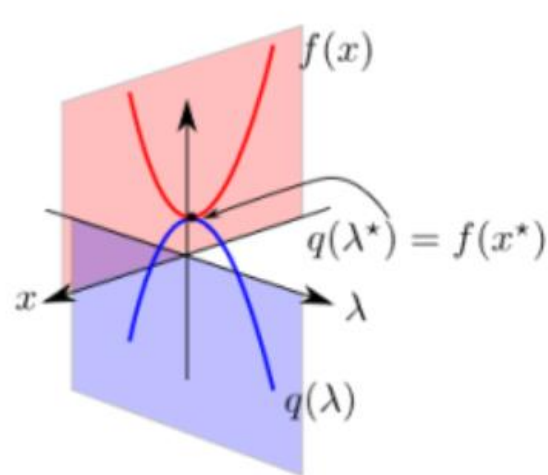
$$g(\lambda) = \inf_x L(x, \lambda) = \inf_x \left(f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) \right)$$

04 | The Lagrangian

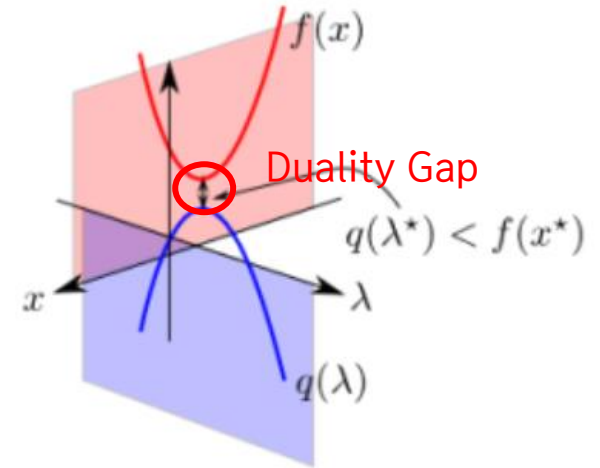
Weak Duality and Strong Duality

1. Weak Duality

$$\begin{aligned} p^* &= \inf_x \sup_{\lambda \succeq 0} \left[f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) \right] \\ &\geq \sup_{\lambda \succeq 0} \inf_x \left[f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) \right] = d^* \end{aligned}$$



strong duality



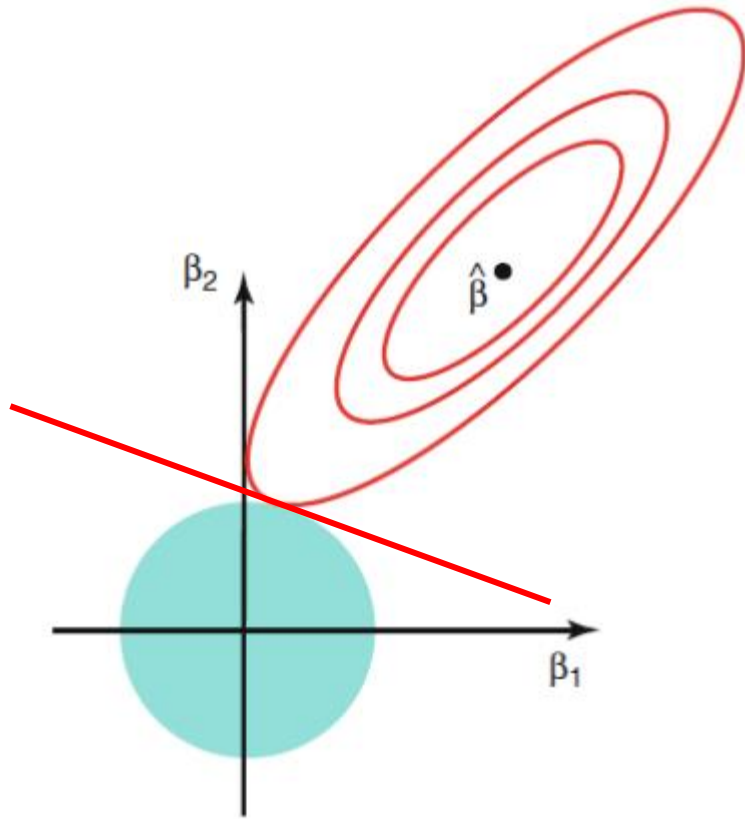
weak duality

2. Strong Duality

$$p^* = d^*$$

05 | Extra Study

The Ridge Regression and The Lagrangian



$$\hat{\beta}^{ridge} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^n (y_i - x_i \beta)^2$$
$$\text{subject to } \sum_{j=1}^p \beta_j^2 \leq t$$

$$\hat{\beta}^{ridge} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^n (y_i - x_i \beta)^2 + \lambda \sum_{j=1}^p \beta_j^2 \right\}$$

→ 제약 식과 목적함수의 접점일 때가
최적화

Q&A

감사합니다.