2. Penalized Regression

KUBIG

Index

1. Ridge Regression
2. LASSO Regression
3. Elastic Net

0. Penalized Regression

- **Why?**

Overfitting 방지

- **How?**

회귀 모델의 **objective function(cost function)**에 **penalty**항을 추가

1. Ridge Regression

- **Ridge 의 objective function = MSE 에 penalty를 부여한 형태**

$$\hat{\beta}_R = arg\min_{\beta}(y - X\beta)'(y - X\beta) + \lambda\|\beta\|_2^2,$$

where $\|\beta\|_2 = \sqrt{\sum_{i=1}^{N}\beta_i^2}$ and λ is the ridge penalty parameter which we have to choose among nonnegative constants.

$$\hat{\beta}_R = arg\min_{\beta}(y - X\beta)'(y - X\beta) \quad \text{s.t.} \quad \|\beta\|_2^2 \le c.$$

1. Ridge Regression

- **Ridge 의 특징**

1. 회귀계수의 제곱합에 penalty 부여 ➡ **L2-norm** 정규화
2. 중요도(설명력)가 적은 변수의 회귀계수 값을 0에 가깝게 감소시킴
3. 다중공선성에 대한 해결책

1. Ridge Regression

다중공선성

⇓

X'X 의 역행렬이 존재하지 않음!

⇓

회귀계수 값을 추정 불가능

The objective function can be written as

$$\begin{aligned} Q_R(\beta) &= (y - X\beta)'(y - X\beta) + \lambda\beta'\beta \\ &= y'y - 2y'X\beta + \beta'(X'X + \lambda I_N)\beta, \end{aligned}$$

and its FOC is given by

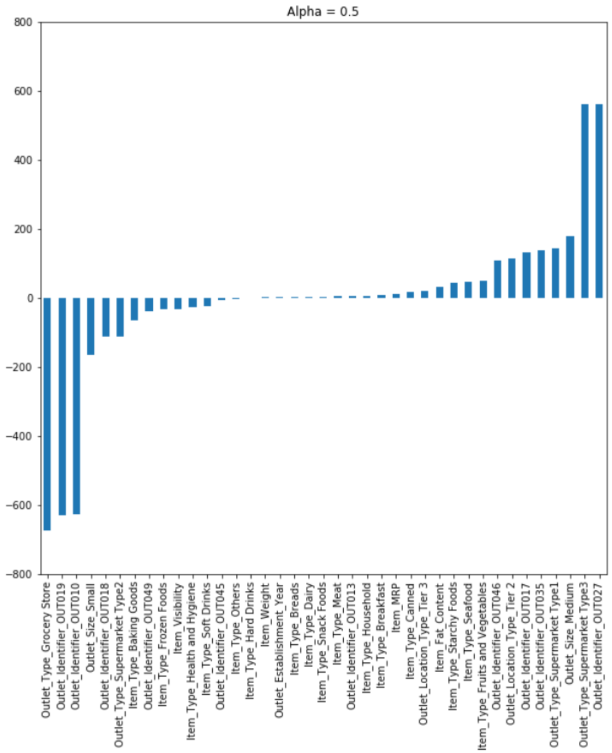
$$\frac{\partial}{\partial\beta} Q_R(\beta) = -2X'y + 2(X'X + \lambda I_N)\beta = 0$$

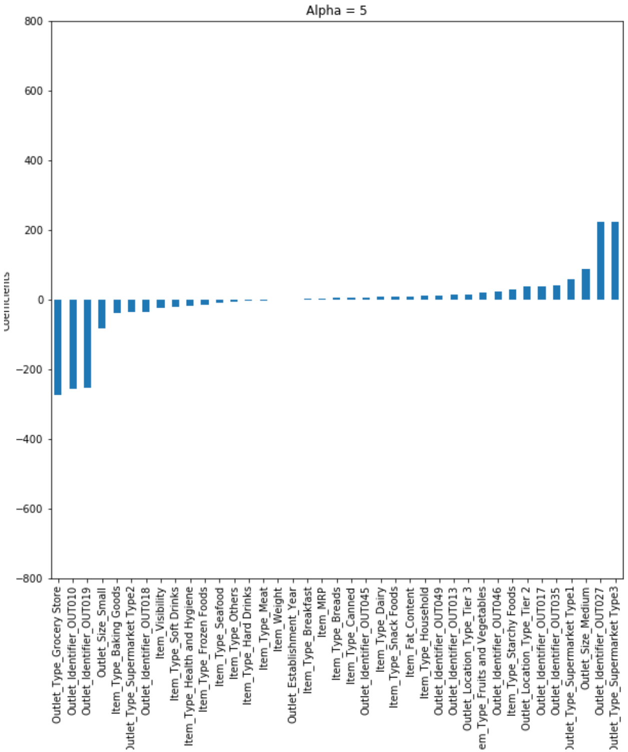
since $\partial(a'x)/\partial x = a$ and $\partial(x'Ax)/\partial x = (A + A')x$. Therefore,

$$\boxed{\hat{\beta}_R = (X'X + \lambda I_N)^{-1}X'y}$$

since $X'X + \lambda I_N$ is always invertible. (Why? Use the spectral decomposition!)

1. Ridge Regression





2. LASSO Regression

- **LASSO**

Least Absolute Shrinkage and Selection Operator

2. LASSO Regression

- **LASSO 의 objective function**

$$\hat{\beta}_L = arg\min_{\beta}(y - X\beta)'(y - X\beta) + \lambda \|\beta\|_1 ,$$

where $\|\beta\| = \sum_{i=1}^{N} |\beta_i|$ and λ is the LASSO penalty parameter which we have to choose among nonnegative constants.

(Regularized Method) The following is an equivalent statement of the ridge regression problem

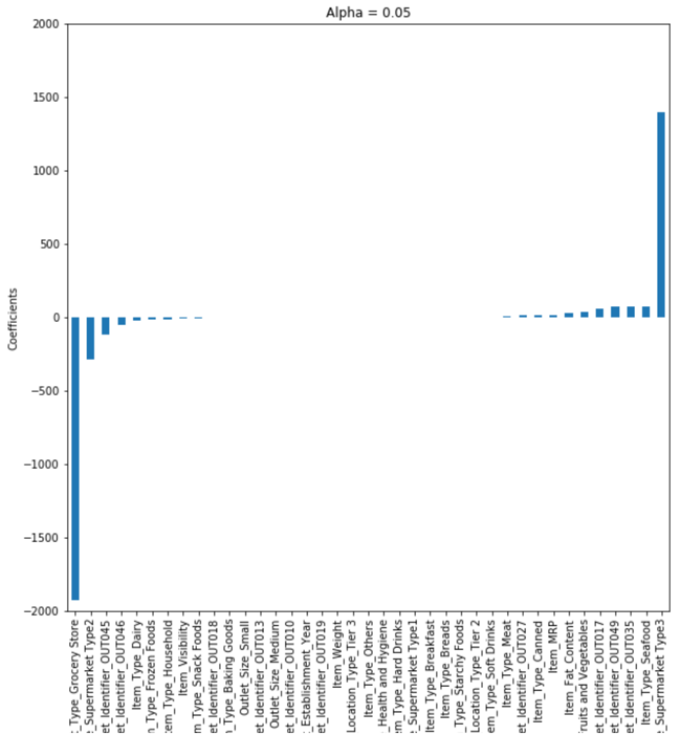
$$\hat{\beta}_R = arg\min_{\beta}(y - X\beta)'(y - X\beta) \quad \text{s.t.} \quad \|\beta\|_1 \leq c.$$

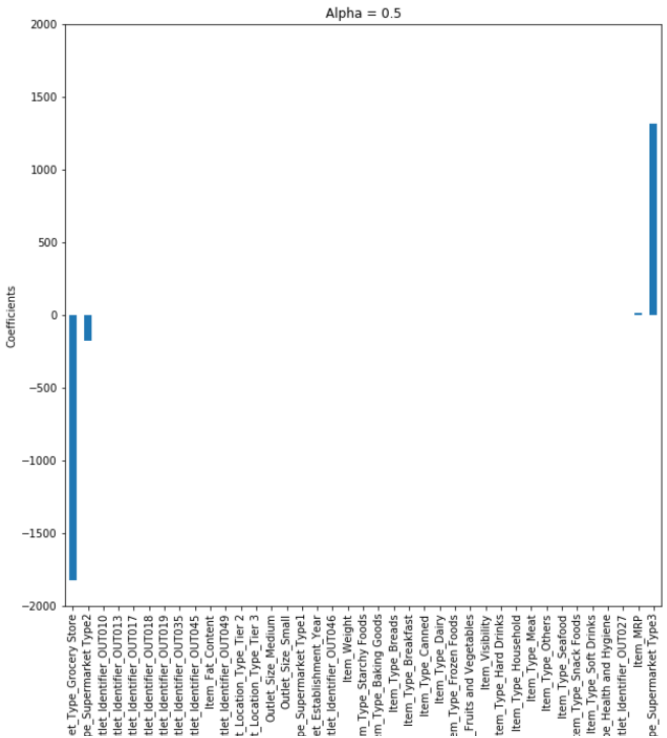
2. LASSO Regression

- **LASSO 의 특징**

1. 회귀계수의 절대값의 합에 penalty 부여 ➡ **L1-norm** 정규화
2. 중요도(설명력)가 적은 변수의 회귀계수 값을 0 까지 줄여 변수 선택의 기능

2. LASSO Regression



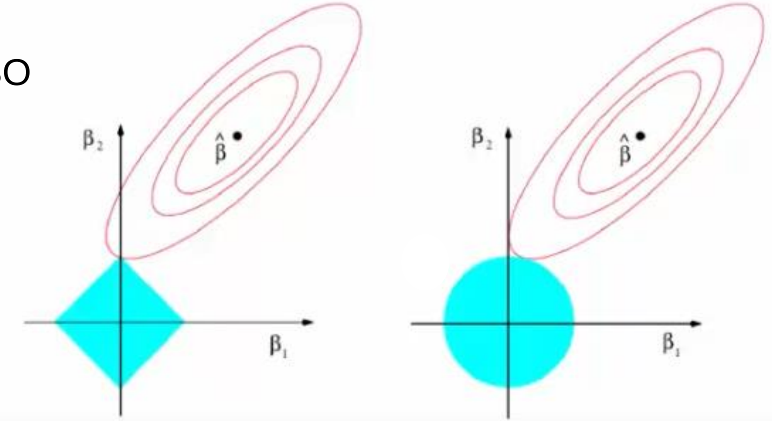


LASSO vs Ridge

- Ridge: minimize the Residual Sum of Square (RSS) with $\sum_{j=1}^{m} \beta_j^2 \leq s$
- Lasso: minimize the Residual Sum of Square (RSS) with $\sum_{j=1}^{m} |\beta_j| \leq s$

LASSO

Ridge



Corner solution

LASSO vs Ridge

	Ridge	Lasso
Regulation	Yes	Yes
Variable selection	No	Yes
Norm	L2	L1
Correlation among variables	Similar coefficients	Select only one of them

변수들 간의 상관관계가 큰 경우, 그 중 하나만 선택함으로써 정보 손실 발생!
(LASSO의 단점)

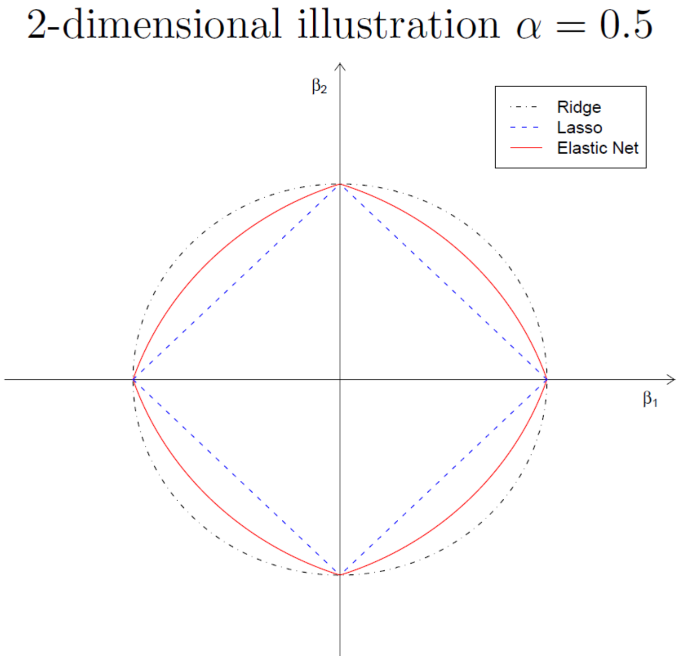
3. Elastic Net

- **Elastic Net 의 objective function = LASSO + Ridge 형태**

$$\hat{\beta}^{enet} = \min_{\beta} \left\{ (y - X\beta)^T (y - X\beta) + \lambda_1 \|\beta\|_1 + \lambda_2 \beta^T \beta \right\}$$

➡ λ1, λ2 의 비율에 따라 Ridge에 가까운지, LASSO에 가까운지 결정됨

3. Elastic Net



$$\alpha = \frac{\lambda 2}{\lambda 1 + \lambda 2}$$

3. Elastic Net

- **Elastic Net 의 특징**

변수들 간의 상관관계가 높은 경우, 그 변수들끼리 그룹을 지어 회귀계수 값을 감소시킴

즉, 하나의 변수만을 선택하는 LASSO와 달리 동시 선택이 가능해짐

KU-BIG

Choosing parameter λ : Cross Validation

λ 는 분석하는 사람이 정해주는 **hyper parameter**

↓

따라서, **λ** 를 변화시키면서 Cross Validation 실시

↓

가장 작은 오차평균(실제값과 예측값의 차이)을 보여주는 **최적의 λ** 를 채택

Thank you