

고려대학교  
빅데이터 연구회

# KU-BIG

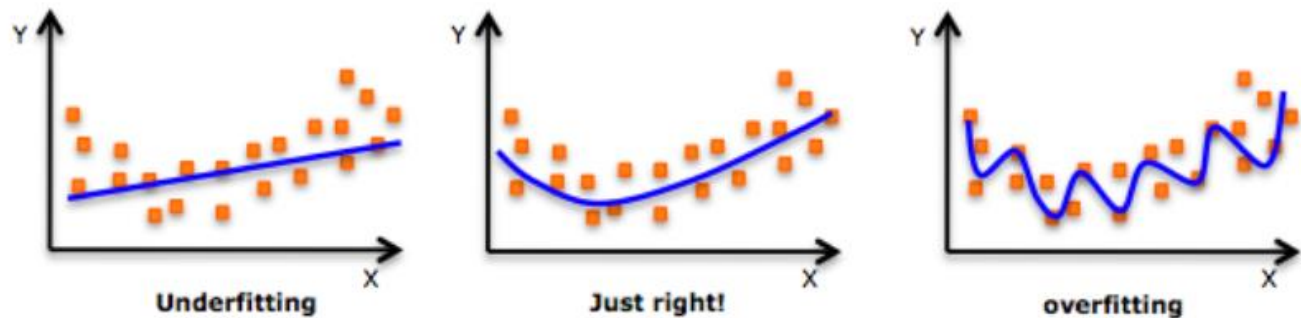
---

Ridge/Lasso/Elastic net

최은혁



# Fitting problems



다중공선선과 같은 문제들 -> 해결방안

# Ridge Regression

$$\hat{\beta}^{Ridge} = \arg \min_{\beta} \left\{ (y - X\beta)^T (y - X\beta) + \lambda \beta^T \beta \right\}$$

최소제곱법 + Penalty

L2 정규화를 이용해 계수 합이 줄어들게 만들어줌

$$\min \left\{ \frac{1}{N} \|y - XB\|^2 \right\} \text{ subject to } \sum \beta^2 < t$$

위 식의 라그랑지 함수 형태

# Ridge Regression

계수 정규화

- 계수들을 줄여주며, 영향이 큰 계수 위주로 줄여줌

# Lasso Regression

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

최소제곱법 + penalty

L1 정규화를 이용해 계수 합이 줄어들게 만들어줌

$$\min \left\{ \frac{1}{N} \|y - XB\|^2 \right\} \text{ subject to } \sum |\beta| < t$$

위 식의 라그랑지 함수 형태

# Lasso Regression

계수 정규화

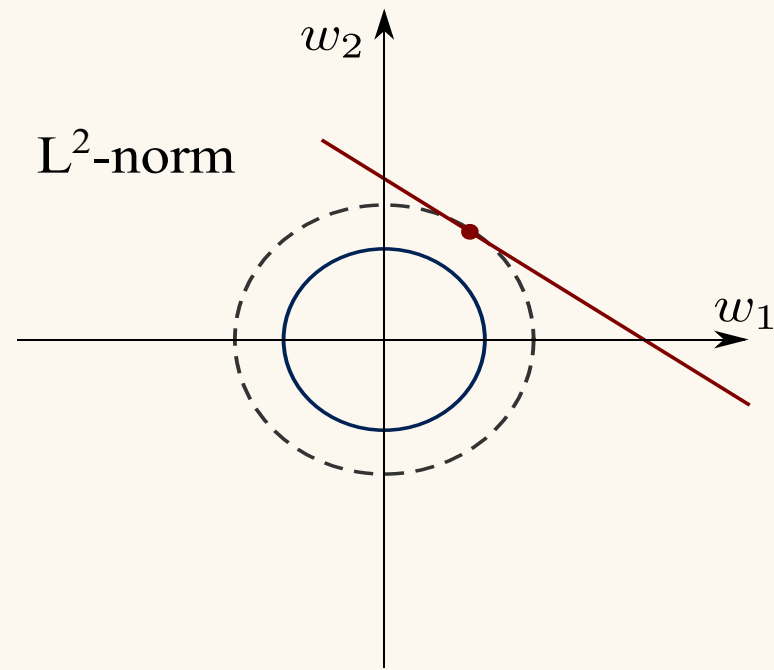
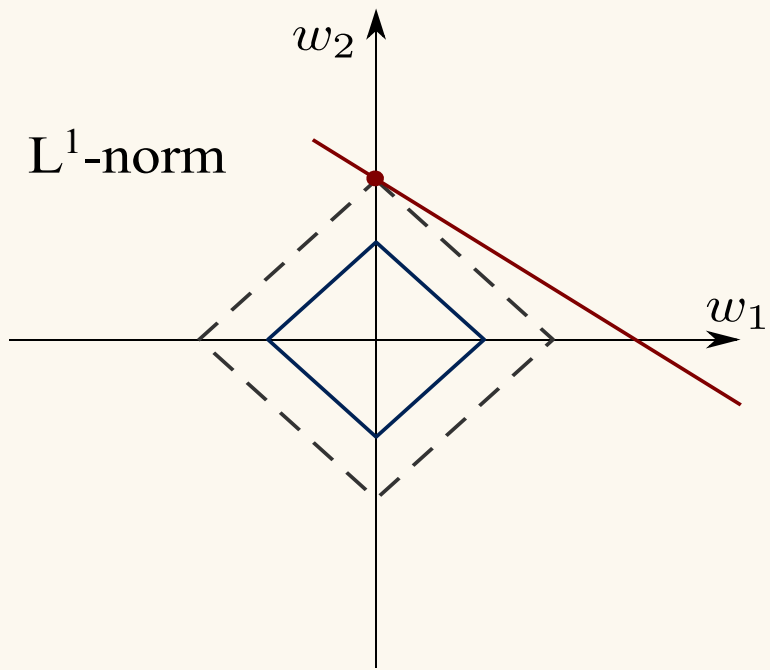
- 계수들을 줄여주며, 비주요 변수 위주로 줄여줌

+

변수 선택

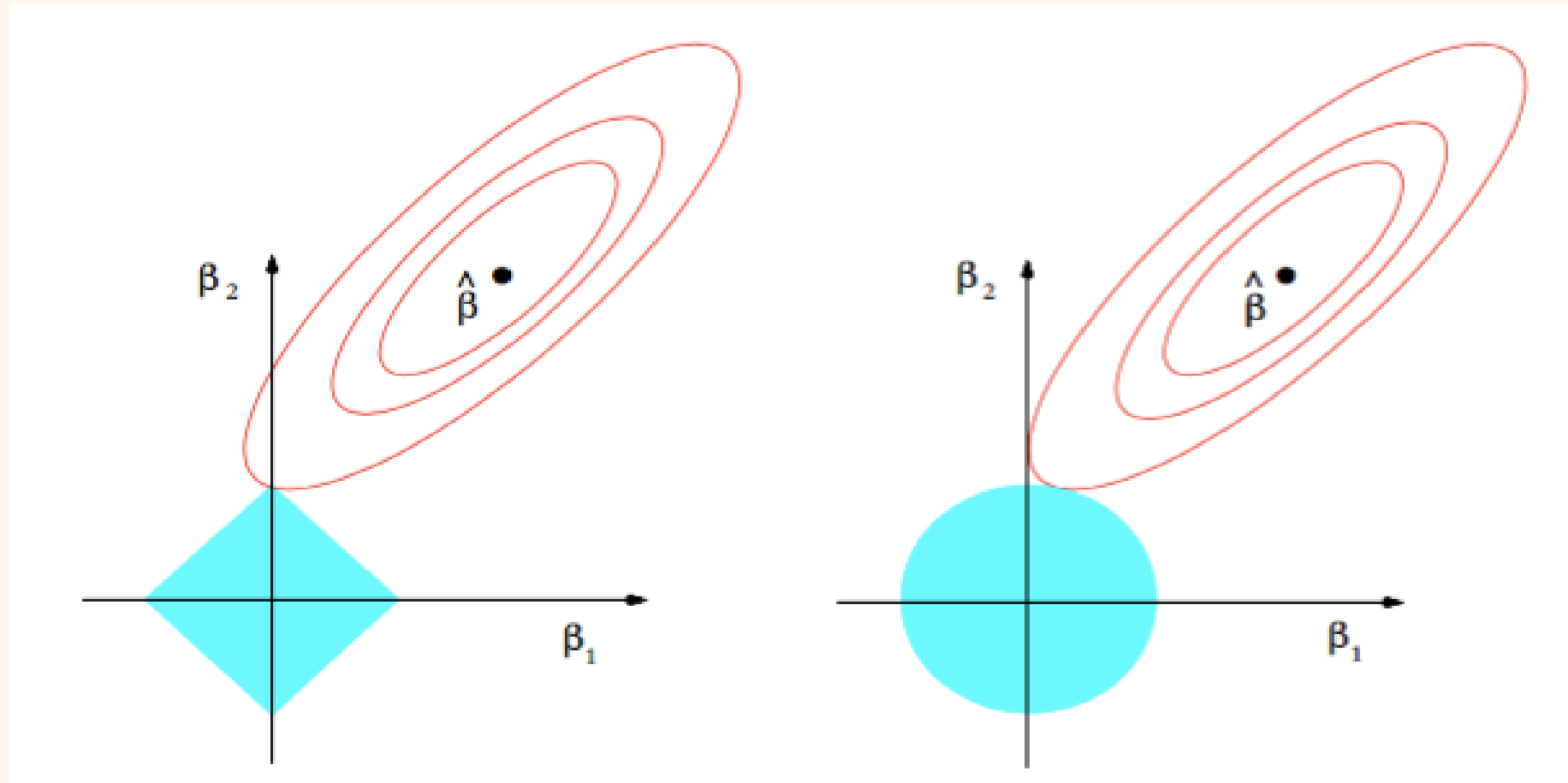
- 비주요 변수를 0까지 줄여서 변수 선택까지 진행해줌

# L1, L2 norm



L1 정규화와 L2 정규화의 형태

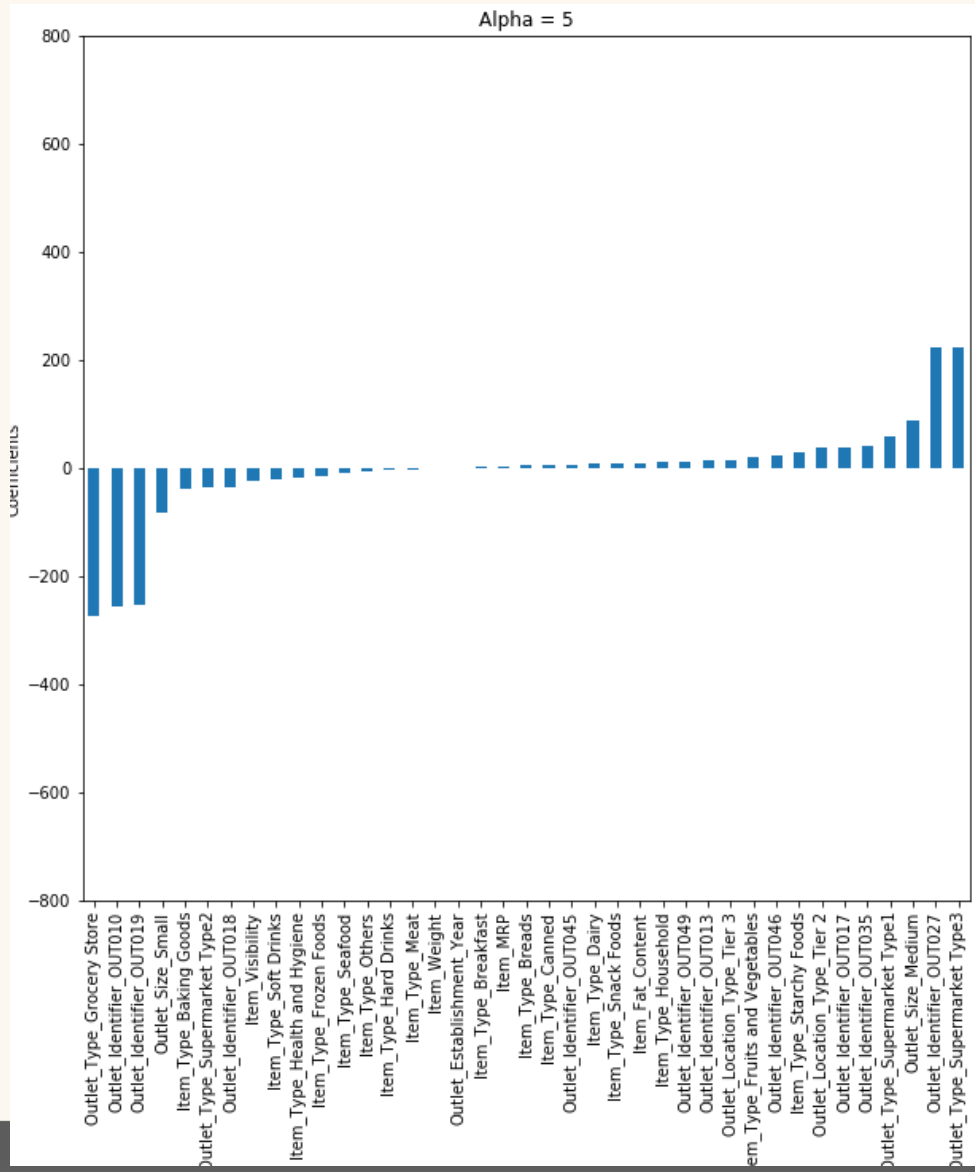
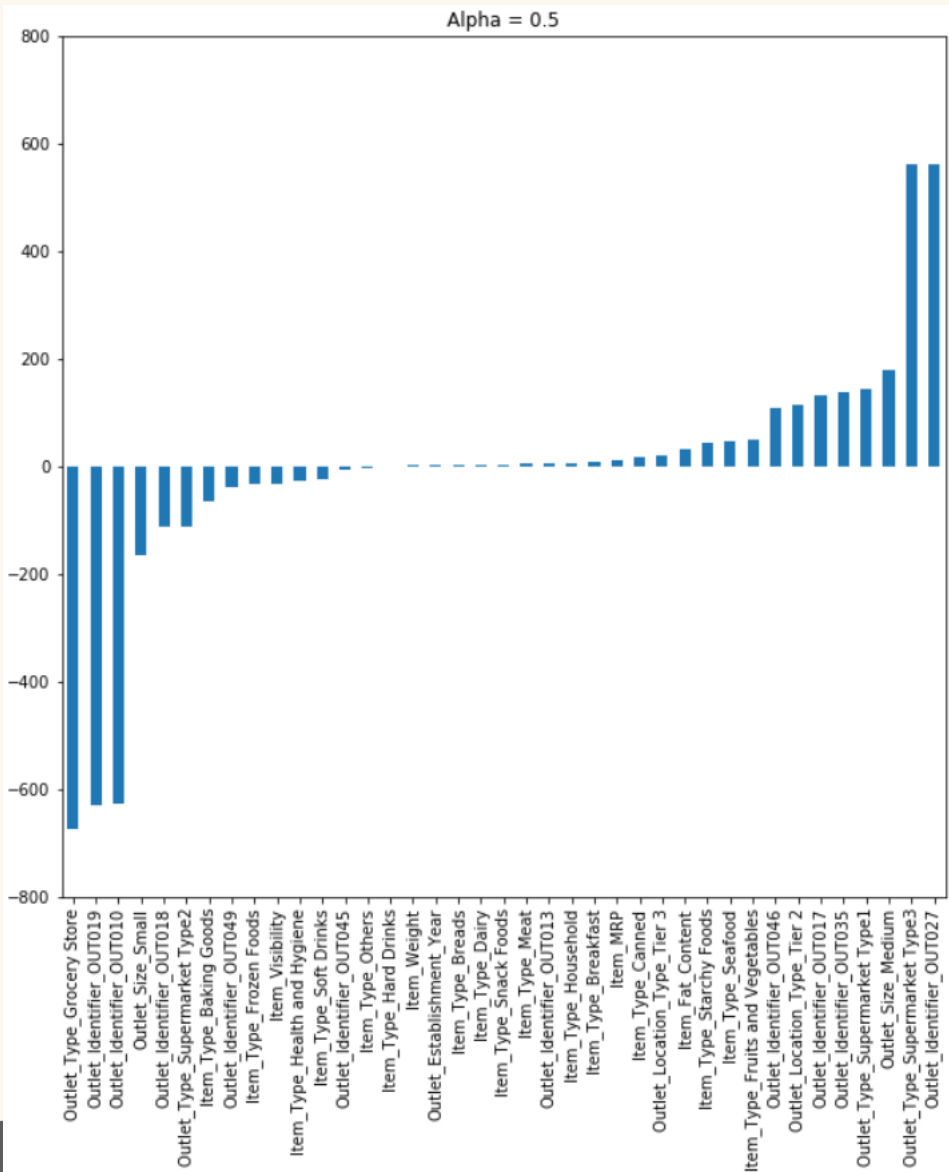
# L1, L2 norm



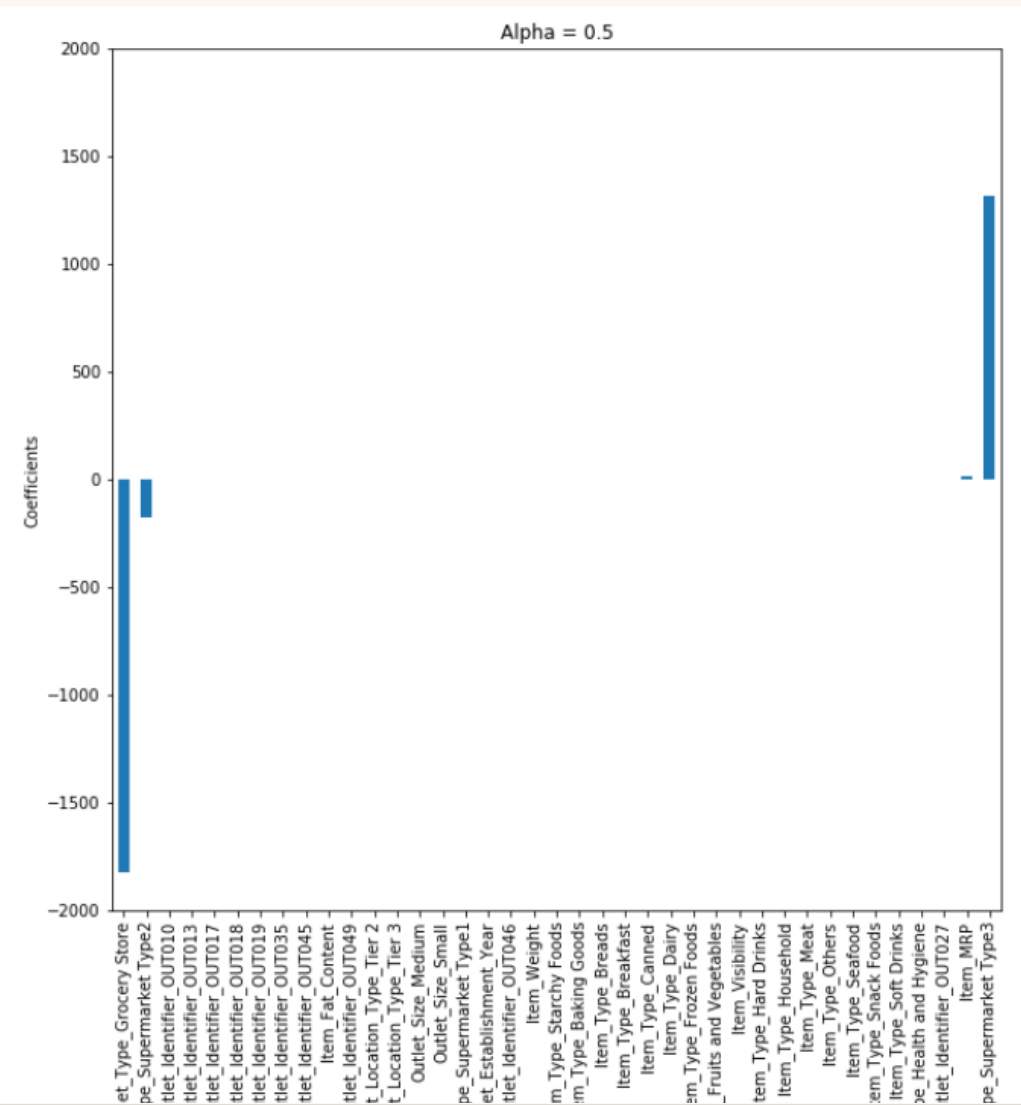
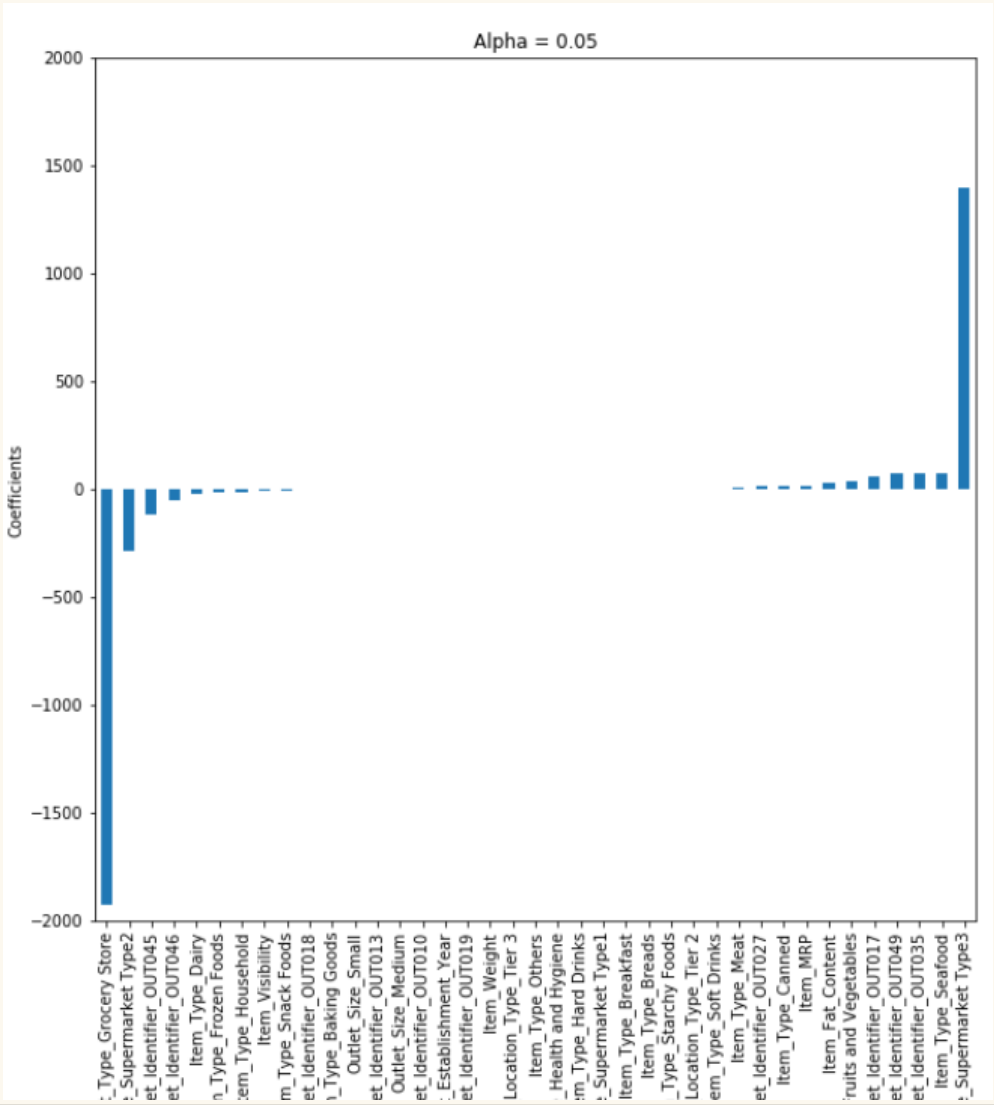


	Ridge	Lasso
Regulation	Yes	Yes
Variable selection	No	Yes
Norm	L2	L1
Correlation among variables	Similar coefficients	Select only one of them

# Ridge



# Lasso



# Importance of scaling

$\hat{\beta}_j$  of variables with small  $\hat{\sigma}_j$ , thus little uncertainty in the estimate, are less penalized;

$\hat{\beta}_j$  of variables with large  $\hat{\sigma}_j$ , thus much uncertainty in the estimate, are heavily penalized.

# Limitation of Lasso (Hui Zou, 2005)

Lasso <- ridge에서 해주지 못하는 변수 선택

변수들 그룹 간의 상관관계가 클 경우: 라쏘는 그룹을 무시하고 그룹의 변수들 중 하나만 남기고 나머지의 계수를 0 <- 정보 손실

$p > n$ 일 경우, 라쏘는  $n$ 개의 변수만을 고르는 경향이 있음

# Elastic Net

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

Lasso의 단점 보완

L1 norm + L2 norm  $\leftarrow \lambda_1, \lambda_2$ 의 비율에 따라

ridge에 가까운지, lasso에 가까운지 결정됨

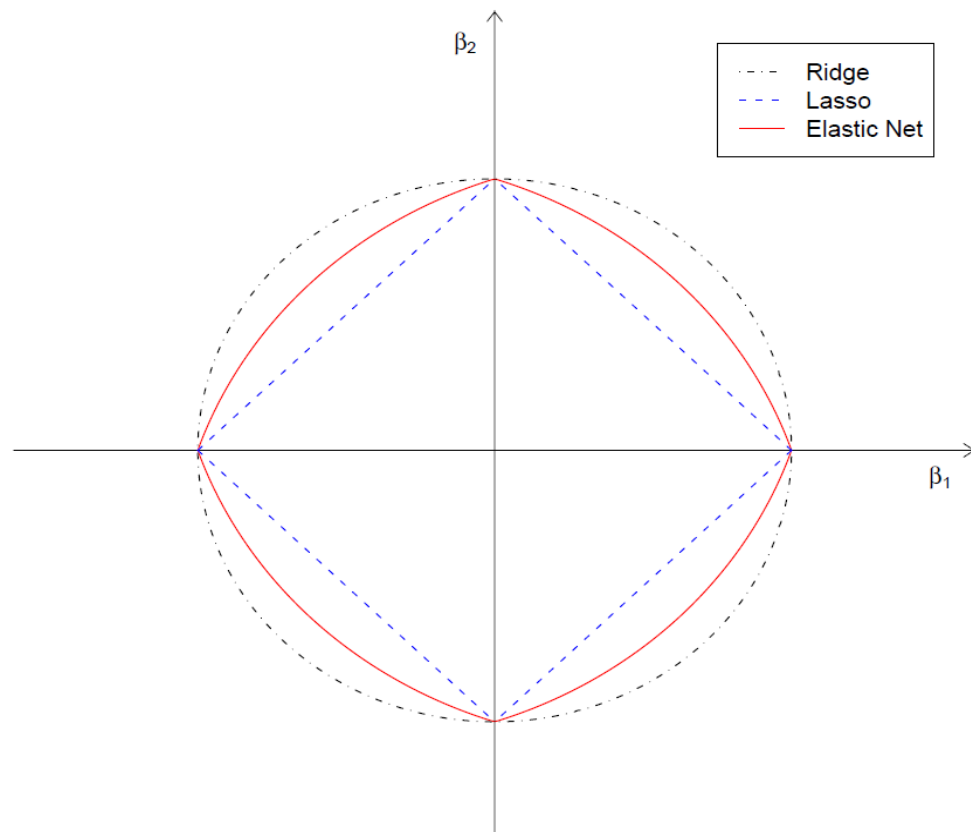
앞서 말한 한계점 해결  $\leftarrow$  regularization + variable selection

+

correlation group : correlation이 큰 그룹의 계수를 통일, 하나가 선택될 경우 모두 선택

# Elastic Net

2-dimensional illustration  $\alpha = 0.5$



$$\alpha = \frac{\lambda_2}{\lambda_1 + \lambda_2}$$

# Choosing tuning parameter $\lambda$

K-fold cross validation을 통해 최적의  $\lambda$  찾기

Elastic net의 경우 parameter가 2개이므로  $\lambda_1$ 을 고정시키고 진행



# Choosing tuning parameter $\lambda$

K-fold cross validation을 통해 최적의  $\lambda$  찾기

Elastic net의 경우 parameter가 2개이므로  $\lambda_1$ 을 고정시키고 진행

# Choosing tuning parameter $\lambda$

K겹 교차 검증(K-fold cross validation) 방법



데이터를 K개로 분할:

1개의 testing set

K-1개의 training set

주어진 값에서

K번 fitting한 후,

K개의 AIC, MSE 등, 오차 평균을 계산

가장 최선의 결과를 낸 값을 채택