CS 307

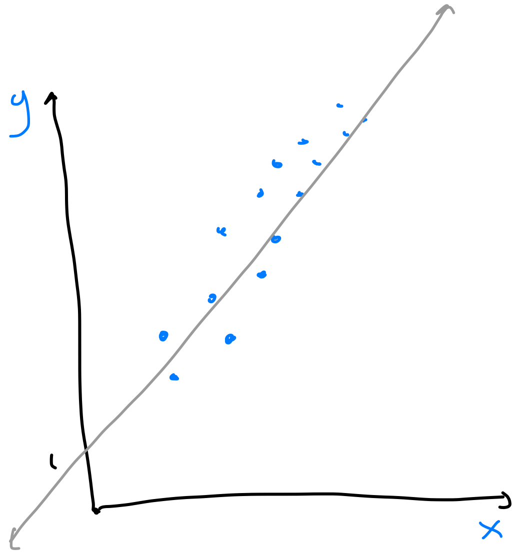
SPRING 2024

DALPIAZ

$\boxed{\text{REGULARIZATION}}$

# Some (Linear) Regression Review

| x | y |
|---|---|
| ⋮ | ⋮ |




## Assume

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

## "Fit"

"Least squares"

$$\min_{\beta_0, \beta_1} \sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 x_i)\right)^2$$

## "Predict"

$$\hat{f}(x) = \hat{\beta}_0 + \hat{\beta}_1 x$$

# "Multiple" Linear Regression

| $y$ | $x_1$ | $x_2$ | ... | $x_p$ |
|---|---|---|---|---|

## Assume

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon$$

## "Fit"

$$\min_{\beta} \sum_{i=1}^{n} \left( y_i - \left( \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} \right) \right)^2$$

$$\min_{\beta} \sum \left( y_i - \left( \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} \right) \right)^2$$

## "Predict"

$$\hat{f}(\underline{x}) = \hat{\beta}_0 + \sum_{j=1}^{p} \hat{\beta}_j x_{ij}$$

# BEST SUBSET SELECTION

$p$ FEATURES, LINEAR REGRESSION

| # FEATURES | # MODELS | MODELS |
|---|---|---|
| $p$ | $1$ | $Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \varepsilon$ |
| $p-1$ | $\binom{p}{2}$ | TOO MANY TO LIST |
| $\vdots$ | $\vdots$ | |
| $2$ | $\binom{p}{2}$ | $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$, $Y = \beta_0 + \beta_1 x_1 + \beta_3 x_3 + \varepsilon$, $\ldots$ |
| $1$ | $p$ | $Y = \beta_0 + \beta_1 x_1 + \varepsilon$, $Y = \beta_0 + \beta_2 x_2 + \varepsilon$, $\ldots$, $Y = \beta_0 + \beta_p x_p + \varepsilon$ |
| $0$ | $1$ | $Y = \beta_0$ |
| | $2^p$ | |

## To save on computation → "Search"

**Step Direction** + **Metric**

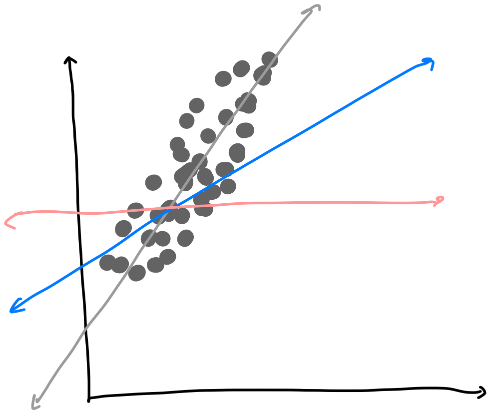
Backwards → AIC → Default in R's step( )

Forwards — BIC

Both — Cp

ETC

Related to LOOCV asymptotically



TRUE MODEL $Y = 2 + 5x + \varepsilon$

LEAST SQUARES

BIAS

$$\min_{\beta_0, \beta_1} \sum_{i=1}^{n} \left( y_i - (\beta_0 + \beta_1 x_i) \right)^2 \longrightarrow \hat{\beta}_1 = 5.2$$

$$\min_{\beta_0, \beta_1} \sum_{i=1}^{n} \left( y_i - (\beta_0 + \beta_1 x_i) \right)^2$$

SUBJECT TO

$$|\beta_1| < 3 \longrightarrow \hat{\beta}_1 = 3$$

$$\min_{\beta_0, \beta_1} \sum_{i=1}^{n} \left( y_i - (\beta_0 + \beta_1 x_i) \right)^2$$

SUBJECT TO

$$|\beta_1| < 0 \longrightarrow \hat{\beta}_1 = 0$$

VAR

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

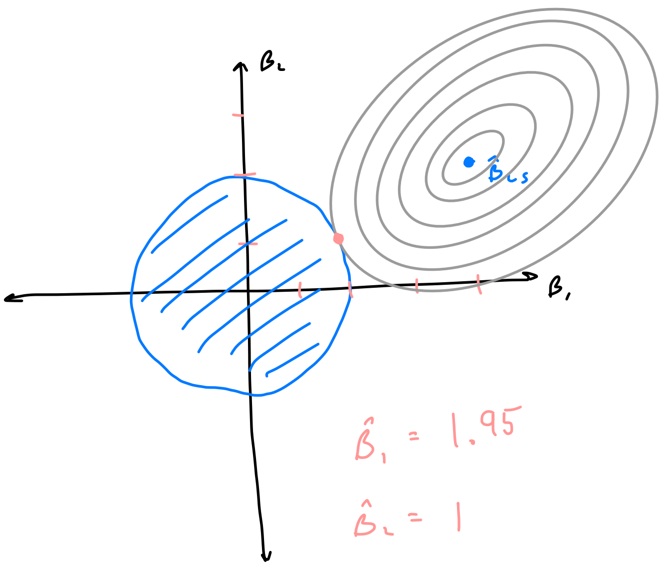
$$\hat{\beta}_1 = 4, \quad \hat{\beta}_2 = 2$$

$$\min_{\beta_0, \beta_1} \sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})\right)^2$$

SUBJECT TO $\beta_1^2 + \beta_2^2 \leq 4$

$$\min_{\beta_0, \beta_1} \sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})\right)^2$$

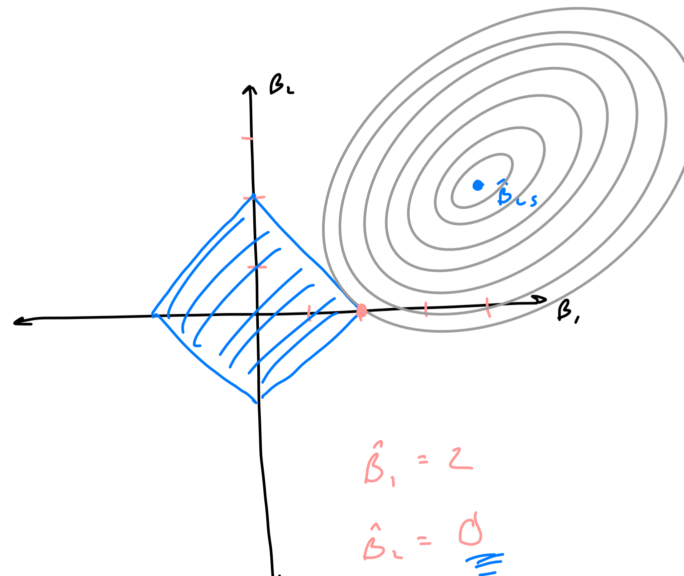
SUBJECT TO $|\beta_1| + |\beta_2| \leq 2$




$\hat{\beta}_1 = 1.95$

$\hat{\beta}_2 = 1$

RIDGE




$\hat{\beta}_1 = 2$

$\hat{\beta}_2 = 0$

LASSO

LS / ERROR

CONSTRAINT

?

$$\min \sum_{i=1}^{n}\left(y_i - \left(\beta_0 + \sum_{j=1}^{p}\beta_j x_{ij}\right)\right)^2 \quad \text{SUBJECT TO} \quad \sum_{j=1}^{p}\beta_j^2 \le s$$

RIDGE

"BUDGET"

$$\min \sum_{i=1}^{n}\left(y_i - \left(\beta_0 + \sum_{j=1}^{p}\beta_j x_{ij}\right)\right)^2 \quad \text{SUBJECT TO} \quad \sum_{j=1}^{p}|\beta_j| \le s$$

LASSO

$$\min \sum_{i=1}^{n}\left(y_i - \left(\beta_0 + \sum_{j=1}^{p}\beta_j x_{ij}\right)\right)^2 \quad \text{SUBJECT TO} \quad \sum_{j=1}^{p} I\left(\beta_j \ne 0\right) \le s$$

BEST SUBJECT SELECTION

RIDGE AND LASSO ARE GREAT WHEN $p$ IS LARGE

LASSO DOES SELECTION!

$$\min \sum_{i=1}^{n} \left( y_i - \left( \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} \right) \right)^2 \quad \text{SUBJECT TO} \quad \sum_{j=1}^{p} |\beta_j| \le s$$

$s = 0 \longrightarrow \beta_0$

$s = \infty \longrightarrow$ OLS

$s \longleftrightarrow \lambda$

$$\min \left[ \underbrace{\sum_{i=1}^{n} \left( y_i - \left( \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} \right) \right)^2}_{\text{ERROR}} + \lambda \underbrace{\sum_{j=1}^{p} |\beta_j|}_{\text{PENALTY}} \right]$$

$\lambda = 0 \longrightarrow$ OLS

$\lambda = \infty \longrightarrow \beta_0$

TUNING

TRADEOFF

AS LS ERROR $\downarrow$, PENALTY $\uparrow$

# A note about scaling

| $y$ | $x$ | $x^*$ |
|---|---|---|
| ⋮ | 900,000 | $-1$ |
| ⋮ | ⋮ | ⋮ |
| ⋮ | 1,000,000 | 0 |
| ⋮ | ⋮ | ⋮ |
| ⋮ | 1,100,000 | 1 |

$$x^* = \frac{x - \bar{x}}{SD[x]}$$

$$Y = \beta_0 + \beta_1 x + \varepsilon \qquad \hat{\beta}_1 = 0.001$$

$$Y = \beta_0 + \beta_1 x^* + \varepsilon \qquad \hat{\beta}_1 = 1000$$

BIG EFFECT ON

$$\sum_{j=1}^{p} \beta_j^2$$

# PENALIZED LOGISTIC REGRESSION

"ERROR" + $\lambda \cdot$ PENALTY

NEGATIVE LOG-LIKELIHOOD

$\ell_1$ "LASSO" $\sum_{i=1}^{p} |\beta_i|$

penalty = "l1"

$\ell_2$ "RIDGE" $\sum_{i=1}^{p} \beta_i^2$

penalty = "l2"