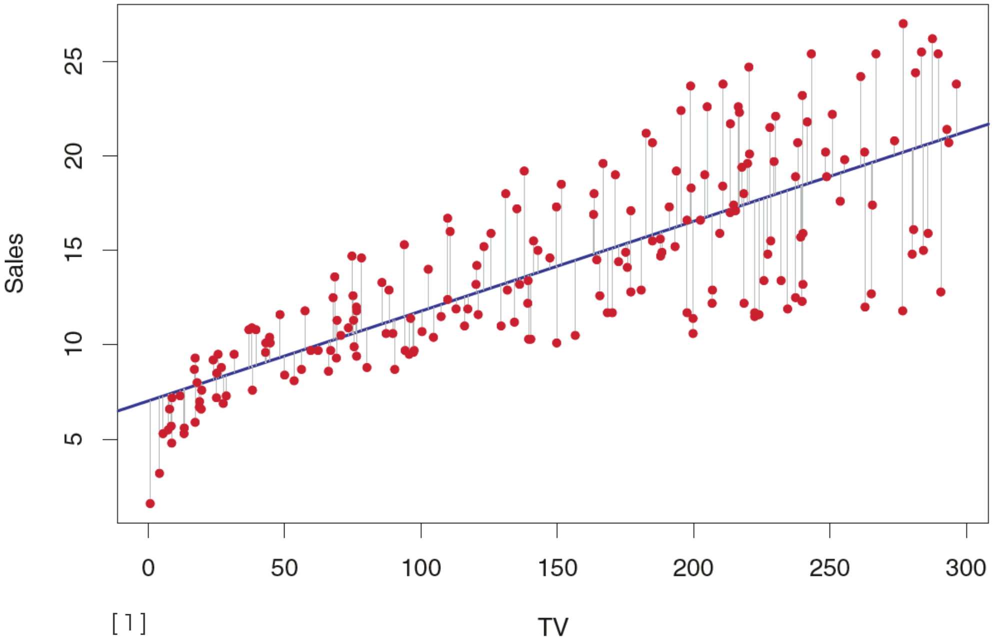
IMMUNE CODING INSTITUTE

Supervised Machine Learning

Module 3

Kahoot!

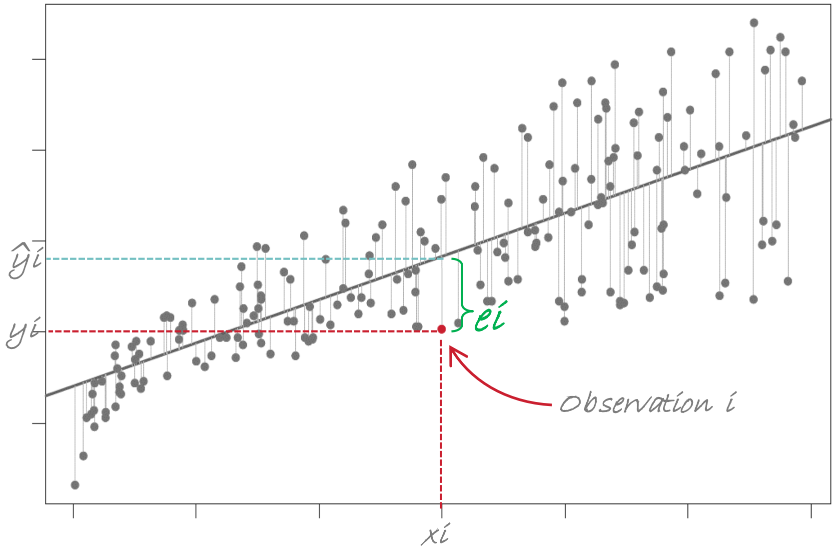
Simple Linear Regression



[1]

- Quantitative predictions
- Single input variable
- $Y \approx \beta_0 + \beta_1 X$
- β_0, β_1:
  - Constant and unknown
  - Coefficients/parameters
- $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$
- $\hat{\beta}_0, \hat{\beta}_1$:
  - Calculated from training data
  - Reduce closeness

Estimate Coefficients



[1]

LEAST SQUARES

$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

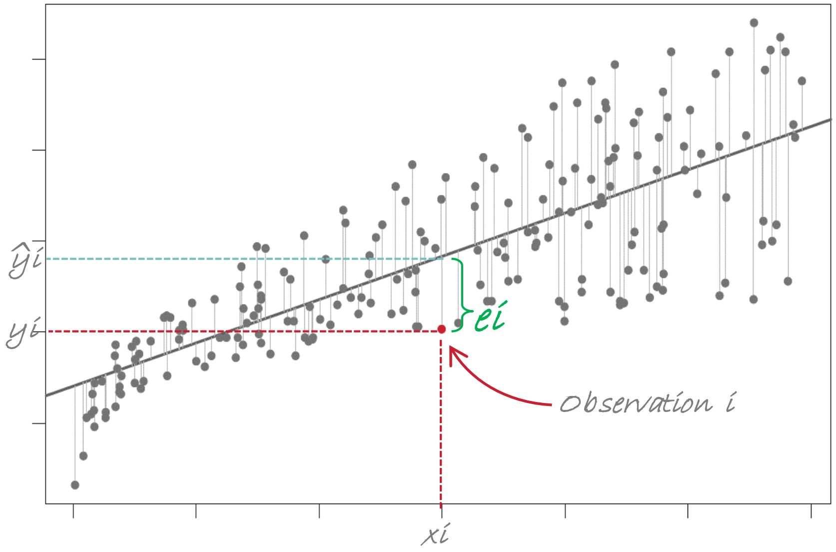
ith residual **>** $e_i = y_i - \hat{y}_i$

Residual Sum of Squares:

$$\mathrm{RSS} = e_1^2 + e_2^2 + \cdots + e_n^2$$
$$= \left(y_1 - \hat{\beta}_0 + \hat{\beta}_1 x_1\right)^2 + \cdots + \left(y_n - \hat{\beta}_0 + \hat{\beta}_1 x_n\right)^2$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Estimate Coefficients



[1]

LEAST SQUARES MATRIX APPROACH

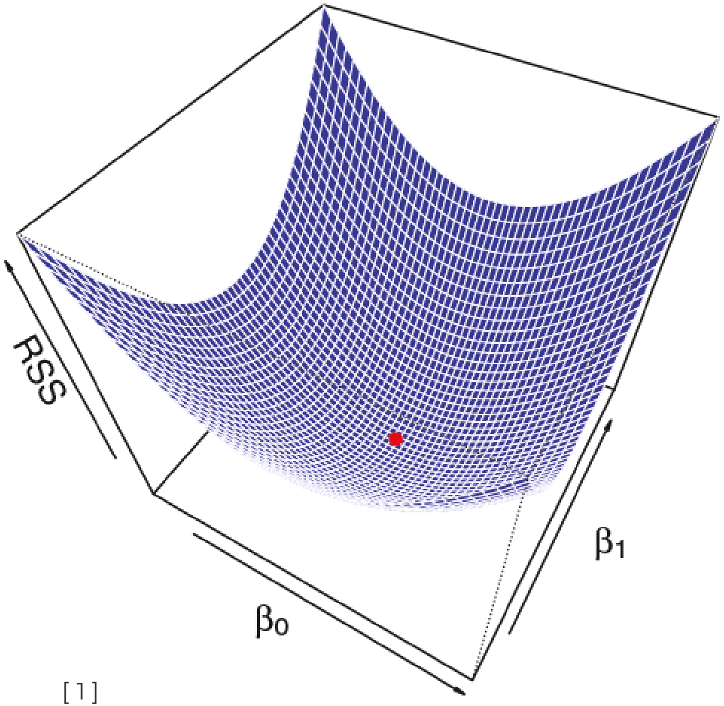
$$\mathbf{X}\boldsymbol{\beta} = \mathbf{y}$$

$$\mathrm{X} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix}, \qquad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}, \qquad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$\hat{\boldsymbol{\beta}} = \left(\mathrm{X}^{\mathsf{T}}\mathrm{X}\right)^{-1}\mathrm{X}^{\mathsf{T}}\mathbf{y}$$

IMMUNE CODING INSTITUTE

Estimate Coefficients

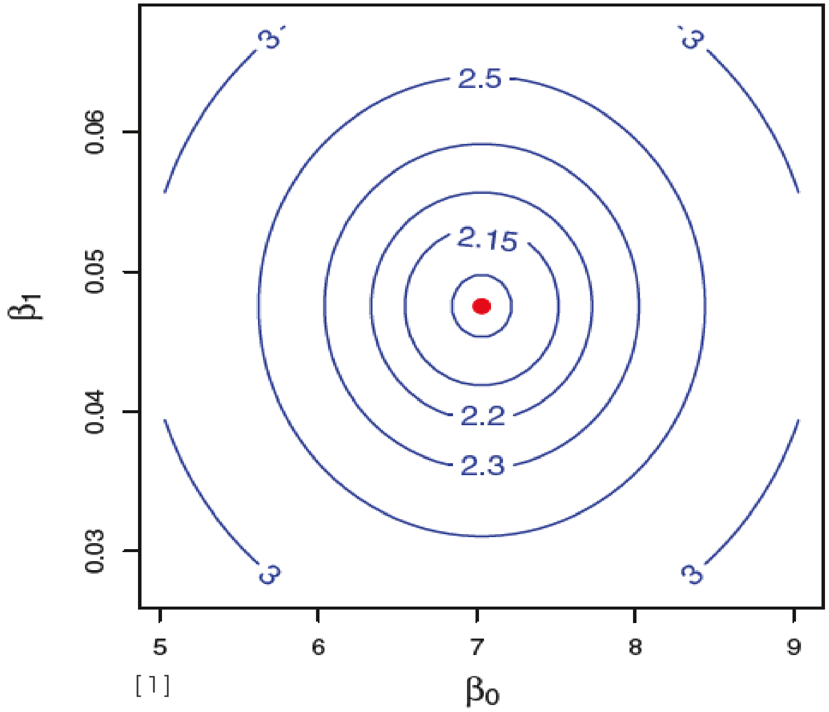


[1]

- This diagram shows how different values for each regression coefficient determine RSS value.
- We can see how there is a single solution for the global minimum of the loss function

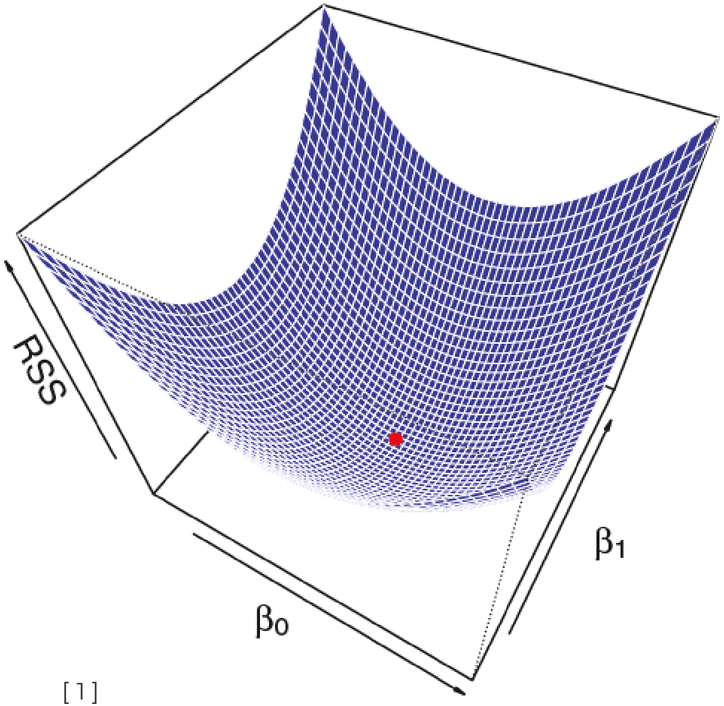
Estimate Coefficients

This diagram shows different ellipses where different values of β lead to same RSS value

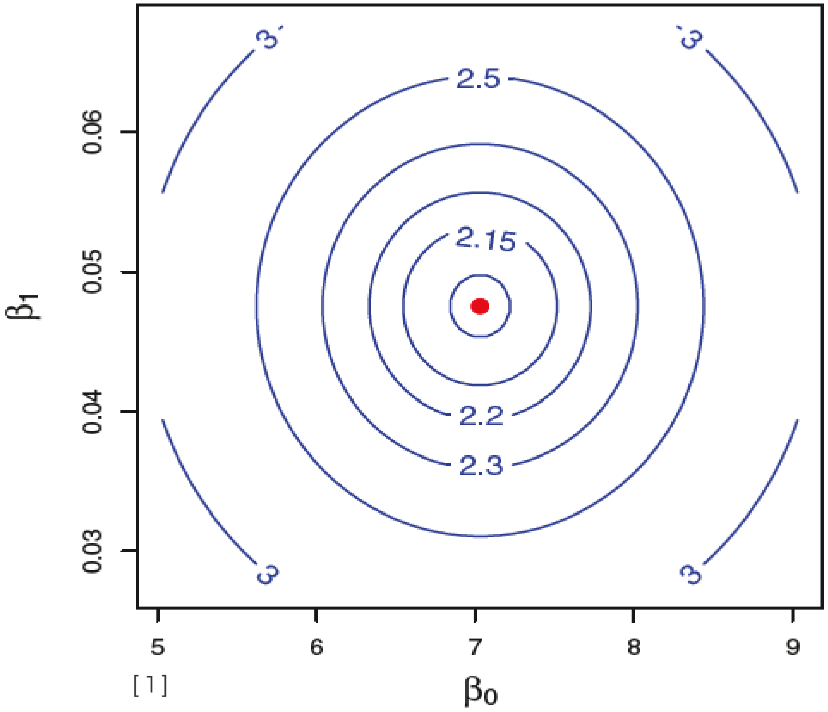


[1]

Estimate Coefficients



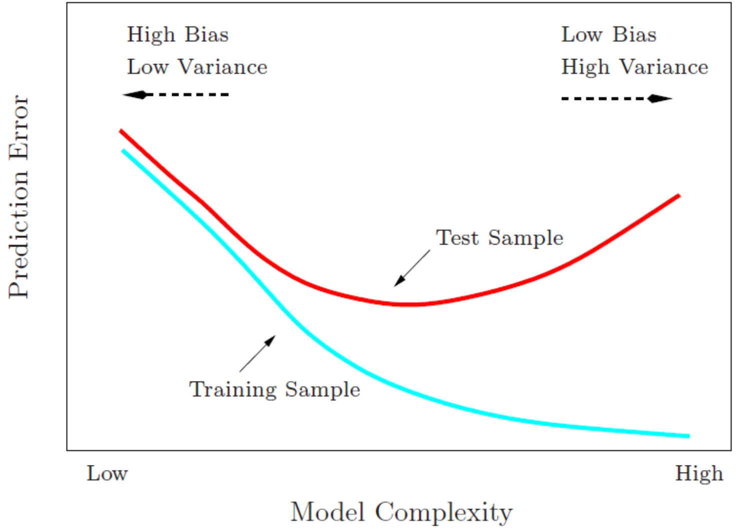
[1]



[1]

Regularization

- Techniques for avoiding overfitting and increasing model interpretability
- Alternative to least squares
- Contrain/regualrize/shrink regression coefficients
- Discourage learning more complex models
- Ridge regression and Lasso Regression



[2]

Ridge Regression

- Similar to OLS but coefficients are slightly minimized by penalty term:

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \underbrace{\lambda\sum_{j=1}^{p}\beta_j^2} = \mathrm{RSS} + \underbrace{\lambda\sum_{j=1}^{p}\beta_j^2}$$

- Tunning parameter λ -> increased for bigger shrinkage penalty

- Predictor standardization:

$$X_{changed} = \frac{X-\mu}{\sigma}$$

IMMUNE CODING INSTITUTE

What would be coefficient value for alpha = 0?

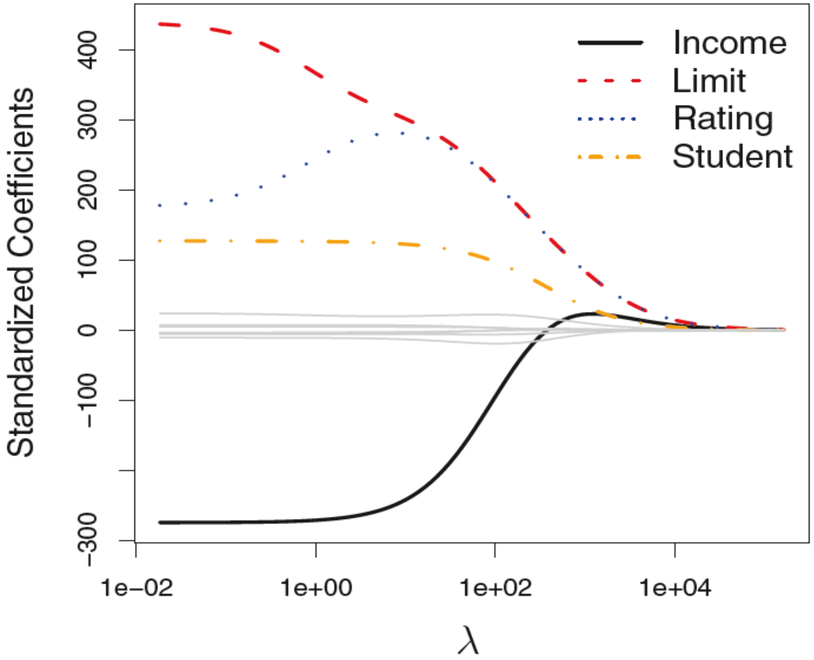
What would be coefficient value for alpha = infinite?

🤔

IMMUNE CODING INSTITUTE

Ridge Regression

- We see how different coefficient values change when we increase alpha value
- As alpha increases, we expect coefficients to decrease
- We may observe sudden increases in some coefficients due to global minimization



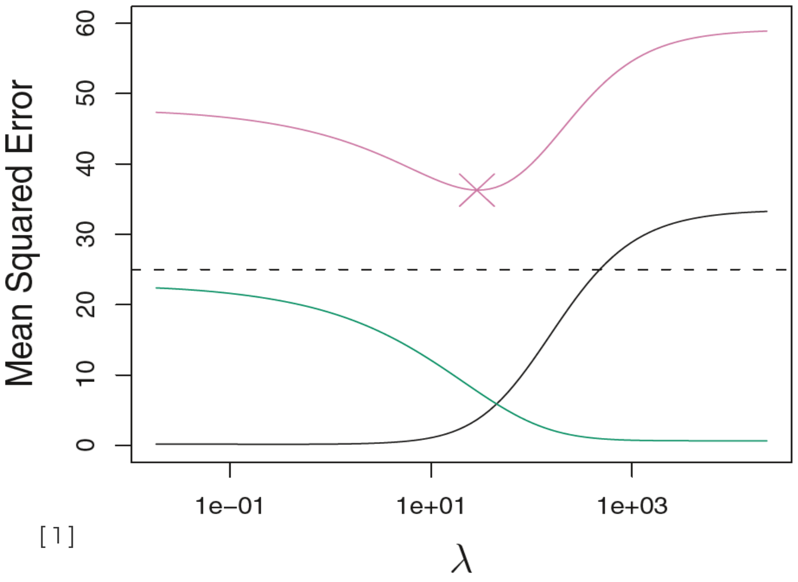
[1]

Ridge Regression

- We see how global MSE(pink), variance MSE (green) and bias MSE(black) vary when we try different Alpha values

- We expect (see graph)

- Our objective is to find the Alpha that corresponds to the mínimum globla MSE



[1]

Lasso Regression

- Ridge regression reduces coefficients, but does not set them to zero -> all predictors are kept in the model

- Lasso coefficients minimize de quantity:

$$\sum_{i=1}^{n}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j| = \mathrm{RSS} + \lambda\sum_{j=1}^{p}|\beta_j|$$

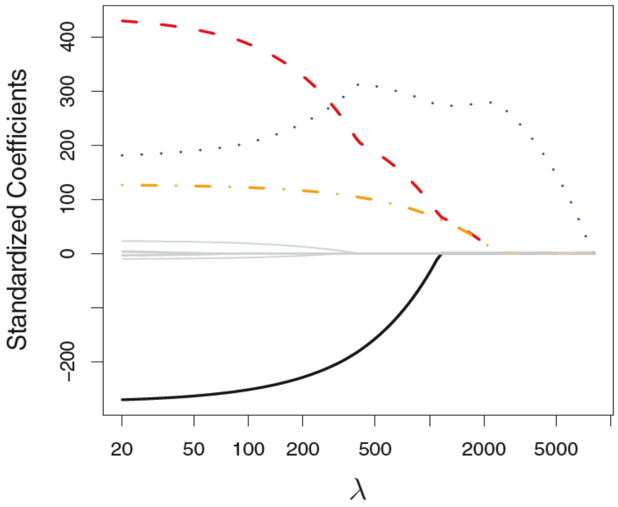
- Lasso forces some coefficients to be exactly equal to zero -> variable selection

- Lasso models are easier to interpret

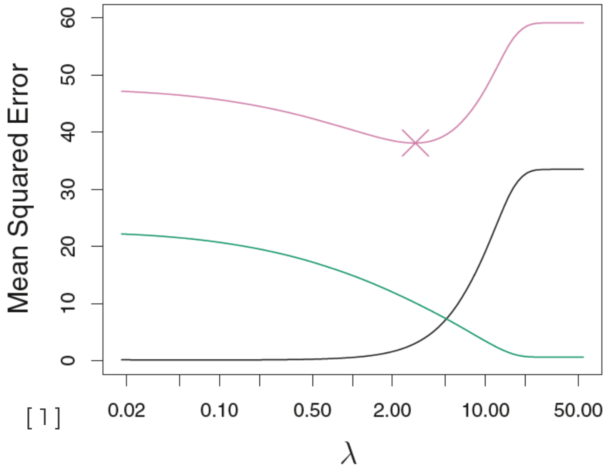
IMMUNE CODING INSTITUTE

Lasso Regression

- As can be now observed, some coefficients yield zero values as Alpha increases



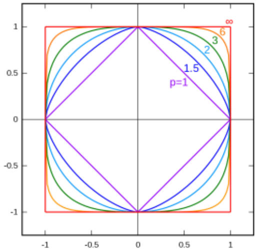
- On the MSE side, we see again the effects of bias-variance tradeoff for different Alpha values



[1]

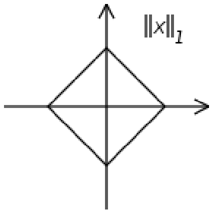
Recall: Norms

$$||x||_p = \left(\sum_i |x_i|^p \right)^{\frac{1}{p}}$$



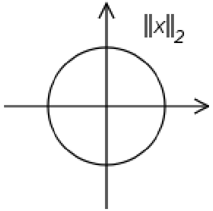
- L^1 **Norm**

  Useful when 0 and non-zero have to be distinguished



- L^2 **Norm**

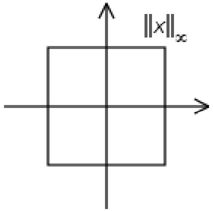
  Called Euclidean norm:
  - Simply the Euclidean distance between the origin and the point x



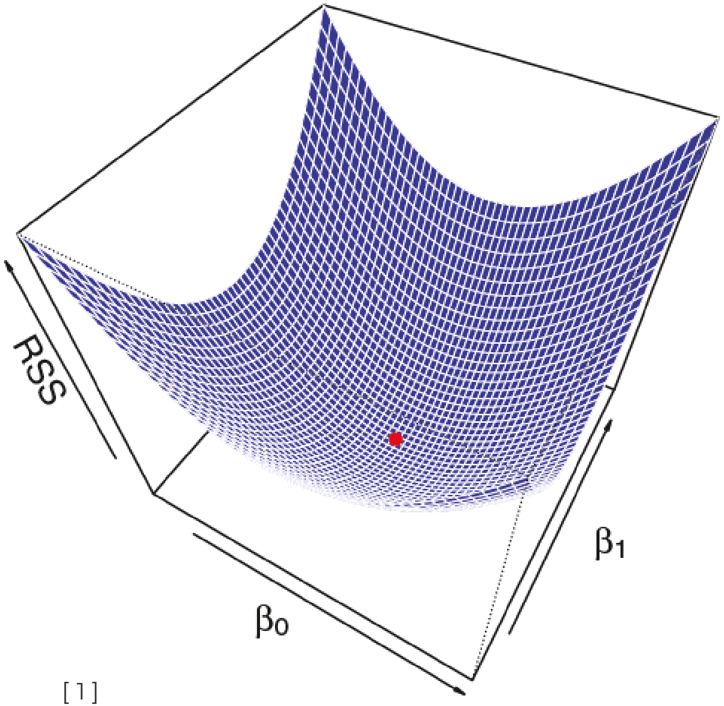
- L^∞ **Norm**

  $$||x||_\infty = max_i |x_i|$$

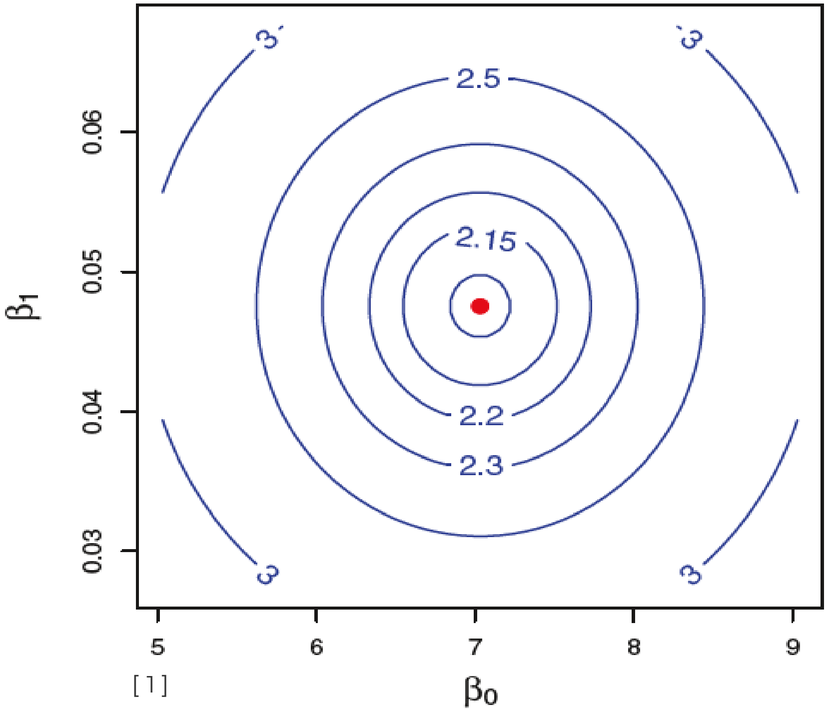
  Called max norm



Recall: Estimate Coefficients



[1]



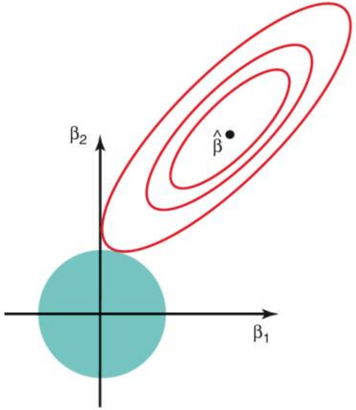
[1]

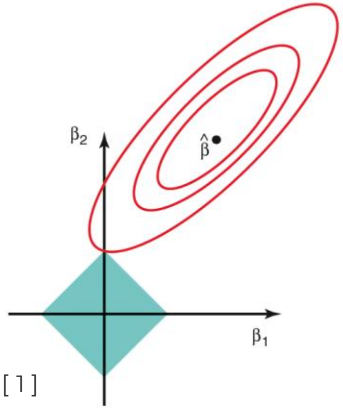
Geometric understanding of regularization

- Different regularization terms lead to different regression solutions

- Ridge regression adds a **L2 regularization term** (circle on the right). This shape make it tough to yield values of beta equal to zero

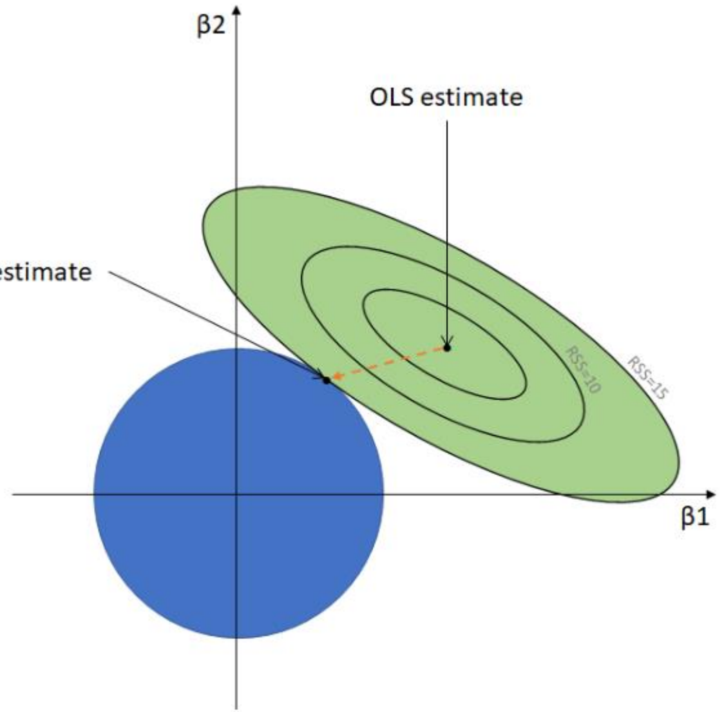
- Lasso adds a **L1 regularization term** (diamond on the right). This sharp shape makes it easier to select beta values equal to zero

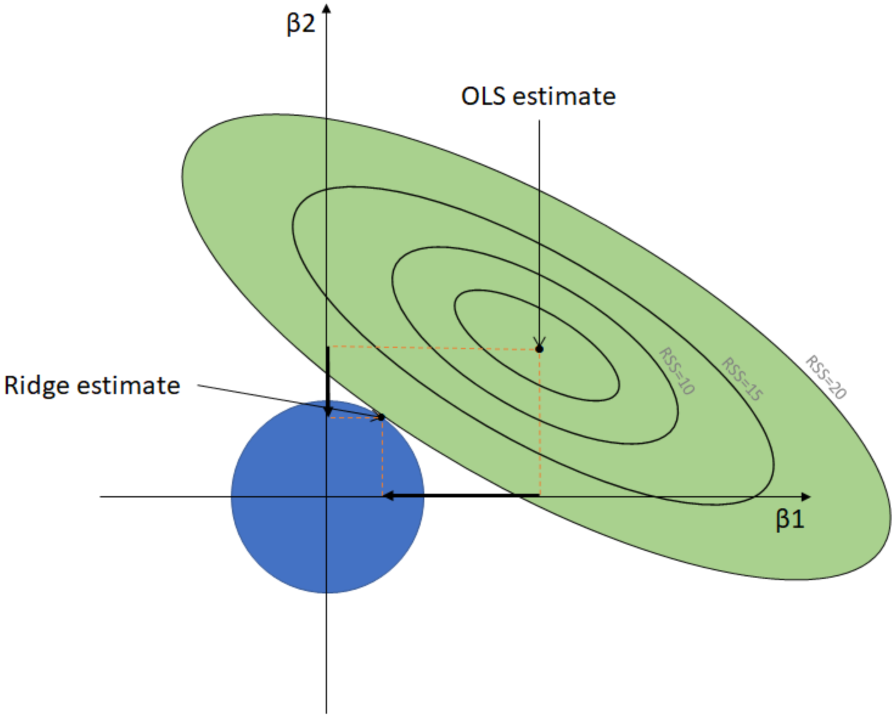




[1]

Geometric understanding of regularization





https://towardsdatascience.com/ridge-regression-for-better-usage-2f19b3a202db

IMMUNE
CODING INSTITUTE

Ridge vs. Lasso

- Lasso has a major advantage on interpretability due to feature selection

- Regarding prediction accuracy, neither ridge regression nor the lasso will universally dominate the other.

- In general:
    - **Lasso**: small number of predictors have substantial coefficients, remaining predictors have coefficients that are very small or that equal zero.

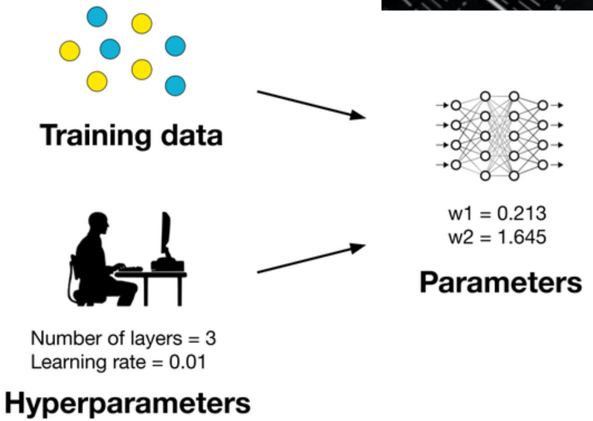
    - **Ridge**: response is a function of many predictors, all with coefficients of roughly equal size.

How would you choose?

Parameters and hyperparameters

- Hyperparameters: parameter whose value is set before the learning process begins.

- Parameters: its value is derived via training.



Instagram @hypertuned.ai

$$\sum_{i=1}^{n}\left(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij}\right)^2+\lambda\sum_{j=1}^{p}|\beta_j| = \mathrm{RSS}+\lambda\sum_{j=1}^{p}|\beta_j|$$

References

[1] G. James, D. Witten, T. Hastie, R. Tibshirani. An Introduction to Statistical Learning with Applications in R. Springer, 2017.

[2] T. Hastie, R. Tibshirani, J. Friedman. The Elements of Statistical: Data Mining, Inference and Prediction. Springer, 2009.

IMMUNE

CODING INSTITUTE