ERASMUS+, OCT 24, 2022

Statistical interpretation of Lasso and Ridge

OLS

$$P(D/\beta) = \prod_{i=0}^{n-1} p(y_i x|\beta)$$

$$P(y_i/\beta) = \prod_{i=0}^{n-1} P(y_i/\beta)$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\left(\frac{(y_i - x_{i*}\beta)^2}{2\sigma^2}\right)}$$

$$x_{i*}\beta = \sum_{j=0}^{p-1} x_{ij}\beta_j$$

Bayes' theorem:

$$P(\beta/D) = ?$$

likelihood prior

$$P(\beta/D) \propto P(D/\beta)P(\beta)$$

Model $p(\beta) = \prod_{j=0}^{p-1} e^{-\beta_j^2/2\tau^2}$

$$P(\beta / D) = \prod_{i=0}^{n-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(y_i - x_{i*}\beta)^2}{2\sigma^2}\right)$$

$$\times \prod_{j=0}^{p-1} \exp\left(-\beta_j^2/2\tau^2\right)$$

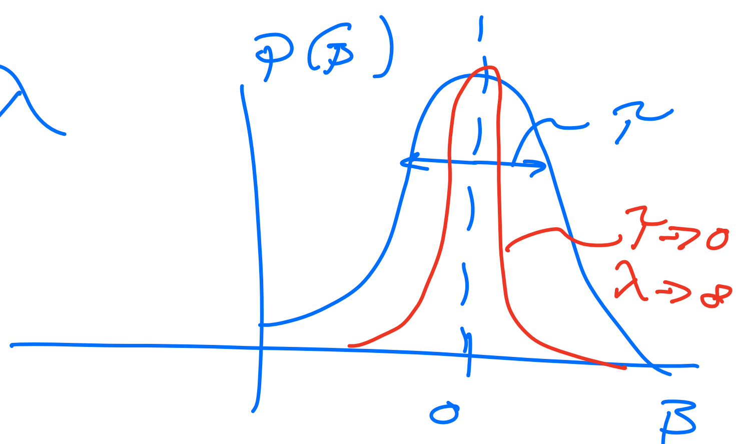
Take negative log

$$-\log P(\beta/D) =$$

$$\frac{n}{2}\log(2\pi\sigma^2) + \frac{\|(y - X\beta)\|_2^2}{2\sigma^2}$$

$$+ \frac{1}{2\tau^2}\|\beta\|_2^2$$

$$\frac{1}{2\tau^2} \to \lambda$$

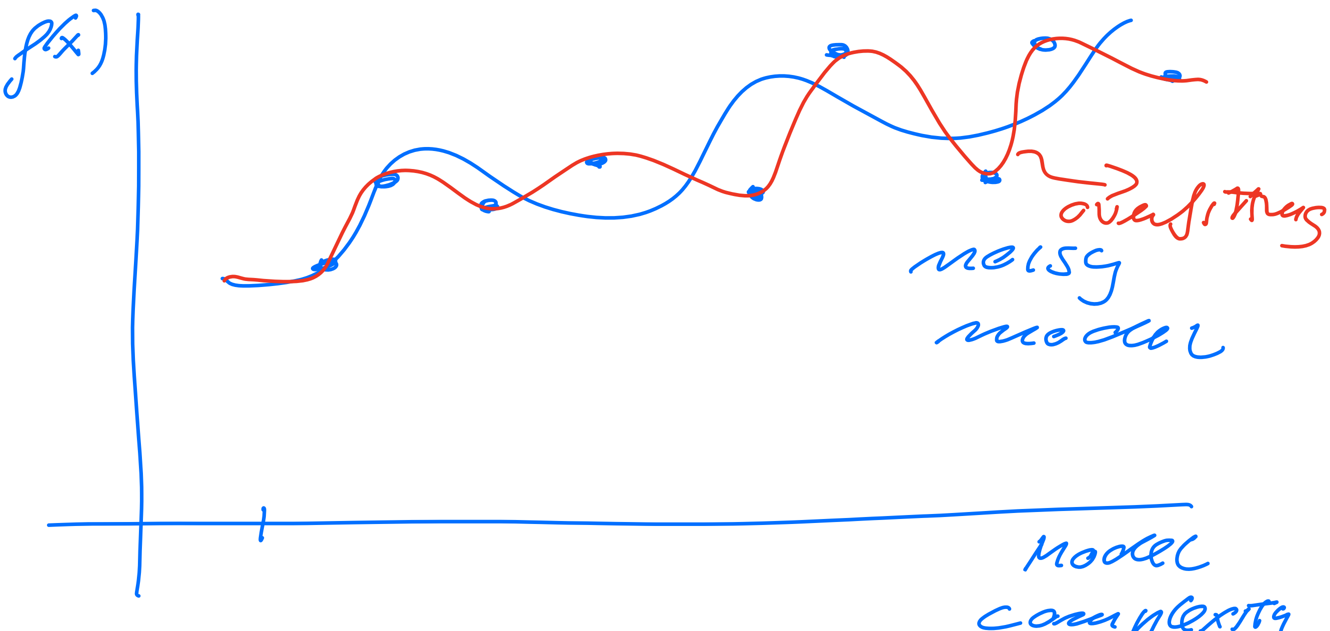


$\tau > 0$

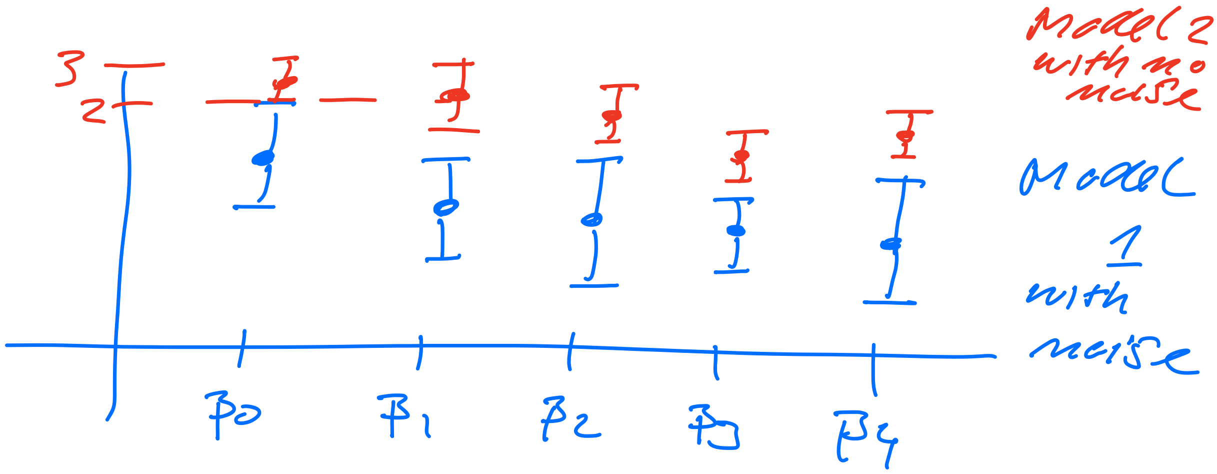
$\lambda > 0$

$$C(\beta)_{Ridge} = \frac{\|(y - X\beta)\|_2^2}{2\sigma^2} + \lambda \|\beta\|_2^2$$

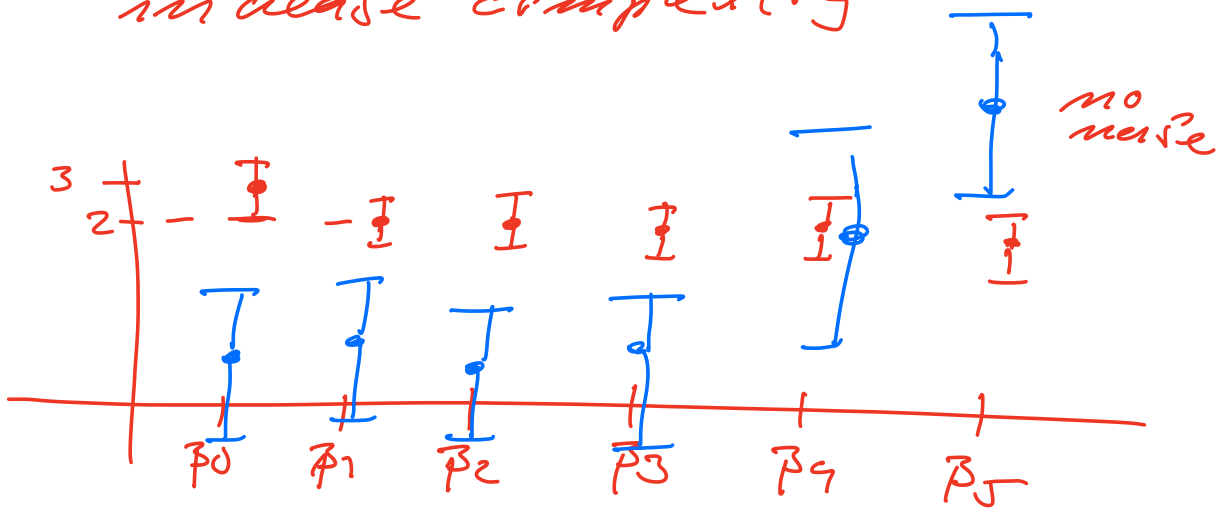
$$p(\beta) \sim N(0, \tau^2)$$



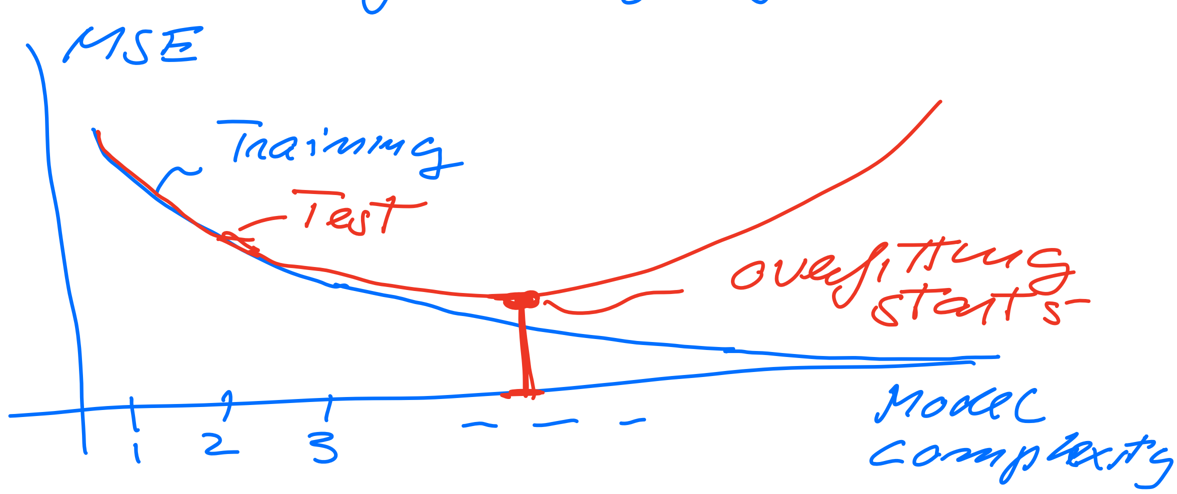
$$f(x_i) = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \beta_4 x_i^4$$

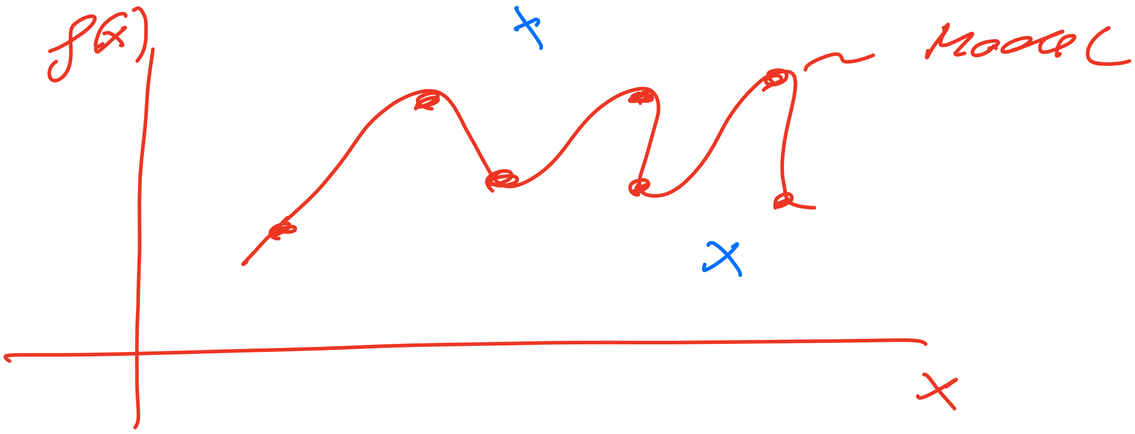


increase complexity



Resampling, underfitting and overfitting of data.





Resampling techniques

- Bootstrap
- cross-validation

Aim: to provide reliable statistical estimates of quantities like mean values and variances, and mean squared errors

mean value

$$\mu = \frac{1}{n}\sum_{i=0}^{n-1} x_i$$

variance

$$\sigma^2 = \frac{1}{n}\sum_{i=0}^{n-1}(x_i-\mu)^2$$

Sample expectation values

$$MSE = \frac{1}{n} \sum (y_i - x_{i*} \beta)^2$$

$$= \mathbb{E}\left((y - x\beta)^2\right)$$

can we, with the limited data we have, give a good estimate (reliable) of various sample expectation values?