Lecture FYS-STK3155/4155, September 9, 2024

FYS-STK 3155/4155, Sept 9

$$y = \underbrace{f(x)}_{\text{non-stochastic}} + \varepsilon$$

$$\varepsilon \sim N(0, \sigma^2)$$

$$\mathbb{E}[x^n] = \int_{x \in D} dx \, x^n \, p_x(x)$$

$$\left(\sum_{x_i \in D} x_i^n \, p_x(x_i) \right)$$

$$\mathbb{E}[x] = \mu_x$$

sample mean

$$\bar{\mu}_x = \frac{1}{n} \sum_{i=0}^{n-1} x_i \neq \mu_x$$

$$\operatorname{var}[x] = \mathbb{E}[x^2] - \mu_x^2$$

$\neq$ sample variance

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

$$MSE = \frac{1}{n} \sum_i (y_i - \tilde{y}_i)^2$$

$$= \mathbb{E}\left(\|y - \tilde{y}\|_2^2\right)$$

Covariance

vectors

$$X = \left[x_0 \; x_1 \; \cdots \; x_{p-1} \right]$$

$$cov(x_j, x_e) = \frac{1}{n} \sum_{i=0}^{n-1} (x_{ij} - \bar{\mu}_{x_j})(x_{ie} - \bar{\mu}_{x_e})$$

$$C[X] = \frac{1}{n} \bar{X}^T X$$

covariance matrix

$$
= \begin{bmatrix}
\mathrm{cov}[x_0, x_0], & \mathrm{cov}[x_0, x_1] & \cdots & \mathrm{cov}[x_0, x_{p-1}] \\
\vdots & \mathrm{cov}[x_1, x_1] & & \\
& & & \\
\mathrm{cov}[x_{p-1}, x_0] & \mathrm{cov}[x_{p-1}, x_1] & \cdots & \mathrm{cov}[x_{p-1}, x_{p-1}]
\end{bmatrix}
$$

$$
= \begin{bmatrix}
\sigma^2[x_0] & & & \mathrm{cov}[x_0, x_{p-1}] \\
\mathrm{cov}[x_1, x_0] & \sigma^2[x_1] & & \\
\vdots & & \ddots & \\
\mathrm{cov}[x_{p-1}, x_0] & \cdots & & \sigma^2[x_{p-1}]
\end{bmatrix}
$$

SVD analysis

$$C[X] = \frac{1}{n} X^T X$$

$$X^T X = V \Sigma^T \Sigma V^T \quad , V$$

$$X^T X V = V \Sigma^T \Sigma$$

$$V = [v_0 \; v_1 \; \ldots \; v_{p-1}]$$

$$X^T X v_i = \sigma_i^2 v_i$$

$$\mathrm{var}[\beta_j]_{OLS} = (X^T X)^{-1}_{jj}$$

$$\mathbb{E}[y_i] = \sum_{j=0}^{p-1} x_{ij}\beta_j = X_{i*}\beta$$

$$\mathrm{var}[y_i] = \sigma^2 \text{ from } \varepsilon \sim N(0,\sigma^2)$$

infer that y follows a normal distribution with the mean value $X_{i*}\beta$ and variance σ^2

$$D = [(x_0, y_0), (x_1, y_1), \ldots (x_{n-1}, y_{n-1})]$$

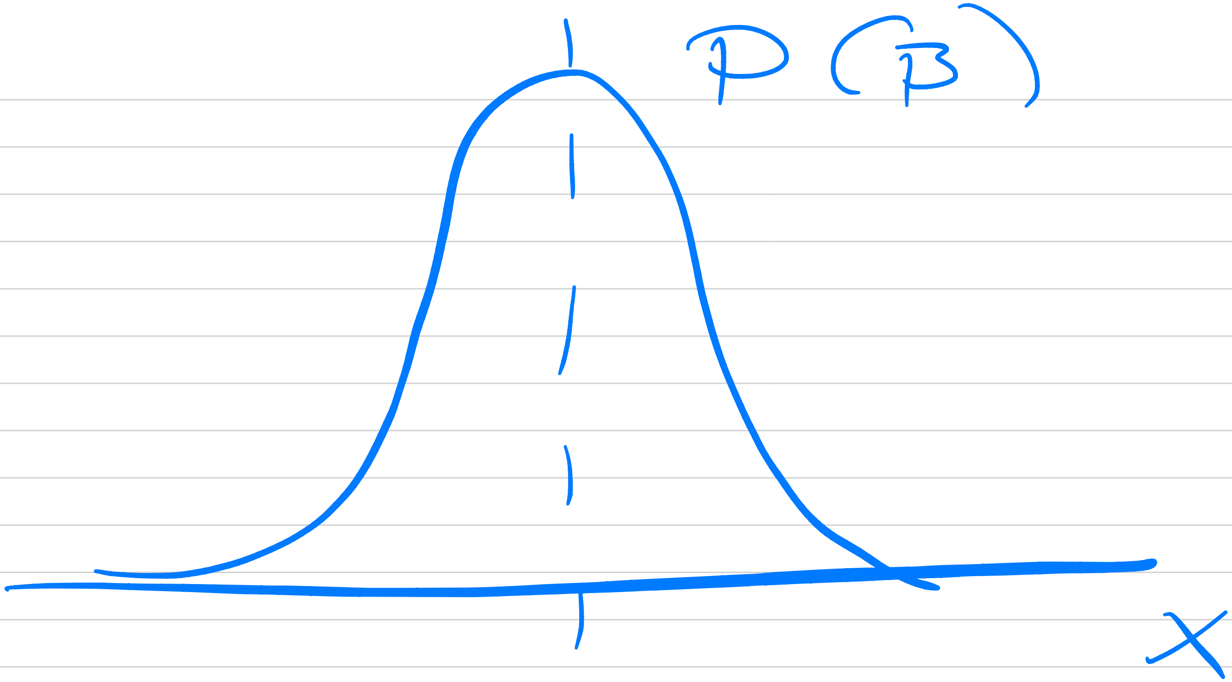
$$P(\underbrace{y_i X}_{\text{all } y_i} | \beta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-(y_i - X_i \beta)^2}{2\sigma^2}\right]$$

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

all y_i are i.i.d. distributed

independent

identically



$$\hat{\beta} = \underset{\beta \in \mathbb{R}^p}{\arg\max} \prod_{i=0}^{n-1} P(y_i, X \mid \beta)$$

MSE

$$\hat{\beta} = \arg\min \frac{1}{n} \|(y - X\beta)\|_2^2$$

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^p}{\arg\max} \log P(D|\beta)$$

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^p}{\arg\min} \left(-\log P(D|\beta)\right)$$

$$-\log \left[\prod_{i=0}^{n-1} P(y_i X|\beta)\right]$$

$$= \frac{n}{2} \log(2\pi\sigma^2) +$$

$$+ \sum_{i=0}^{n-1} \frac{(y_i - x_i * \beta)^2}{2\sigma^2}$$

$$\frac{\partial}{\partial \beta}\left[- \log p(D|\beta) \right]$$

$$\Rightarrow \quad \hat{\beta} = (X^T X)^{-1} X^T y$$

standard OLS

Ridge

$$c(\beta) = \frac{1}{n} \| (y - X\beta) \|_2^2 + \lambda \|\beta\|_2^2$$

Bayes' theorem ?

$$\underbrace{P(\beta|D)}_{\text{posterior}} \propto \underbrace{P(D|\beta)}_{\text{likelihood}} \underbrace{P(\beta)}_{\text{prior}}$$

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

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

$$\times \prod_{j=0}^{p-1} \exp\left[-\frac{\beta_j^2}{2\tau^2}\right]$$

$$\hat{\beta} = \arg\min_{\beta} \left(-\log P(\beta|D)\right)$$

(leave out constant not depending on β)

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

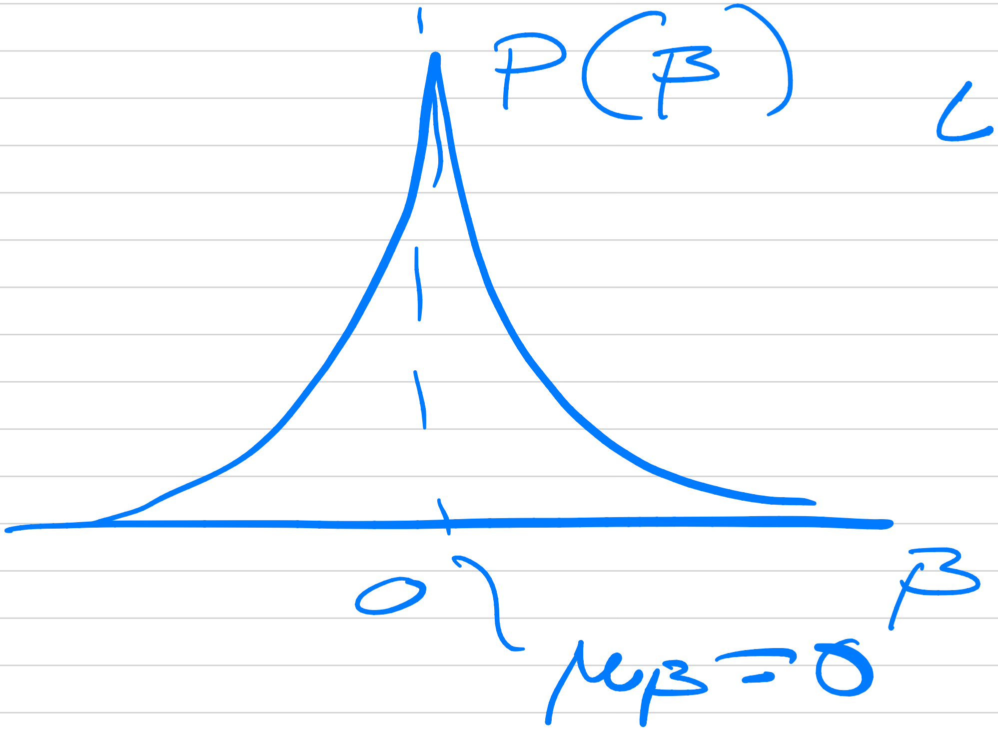
$$\frac{\| (y - X\beta) \|_2^2}{2\sigma^2} + \underbrace{\frac{1}{2\tau^2}}_{\lambda} \|\beta\|_2^2$$

$\lambda \propto$ inverse variance of β

Lasso:

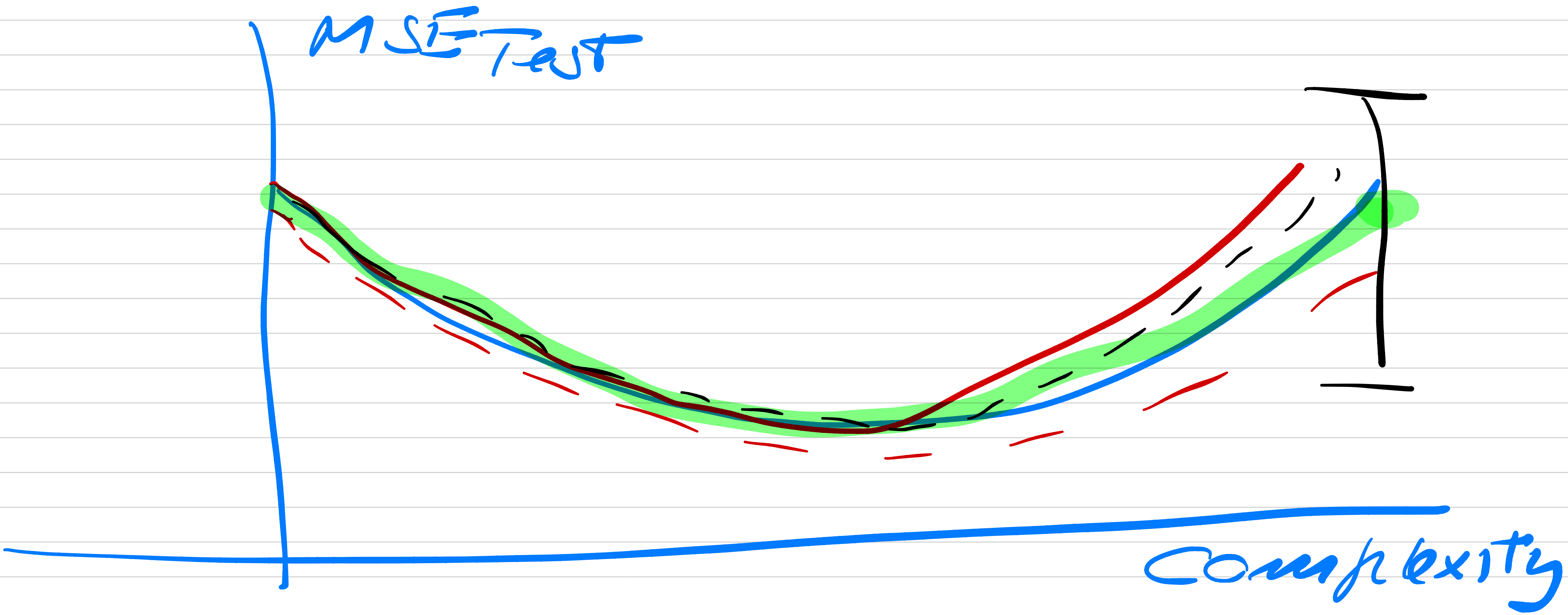
$$P(\beta_j) = \exp\left(-\frac{|\beta_j|}{\tau}\right)$$

Laplace $P(\beta)$



Resampling techniques

- Bootstrap
- k-Fold cross validation



Bootstrap

algorithm

$$X = [x_0\ x_1\ x_2\ \dots\ x_{n-1}]$$

$$\left(\mathbb{E}[X] = \mu_X \neq \frac{1}{n}\sum_{i=1}^{n-1} x_i = \bar{\mu}_X \right)$$

For $i = 1, B$

$$X_i = [x_0^*\ x_1^*\ x_2^*\ \dots\ x_{n-1}^*]$$

with replacement.

Example $X = [1, 3, 7, 10]$

$X^{*} = [1, 7, 7, 10]$

$X^{**} = [7, 10, 3, 3]$

$\vdots$