Lecture FYS-STK3155/4155, September 2, 2024

FYS-STK3155/4155 Sept 2

SVD-decomposition

$$X = U \Sigma V^T \in \mathbb{R}^{n \times p}$$

$$U^T U = U U^T = \mathbb{1} \in \mathbb{R}^{n \times n}$$

$$V V^T = V^T V = \mathbb{1} \in \mathbb{R}^{p \times p}$$

$$U = [u_0 \; u_1 \; \dots \; u_{n-1}]$$

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

$$u_i^T u_j = \delta_{ij} \quad \wedge \quad v_i^T v_j = \delta_{ij}$$

$$\Sigma = \begin{bmatrix} \sigma_0 & & & \\ & \sigma_1 & & 0 \\ & & \ddots & \\ 0 & & & \sigma_{p-1} \\ & & & 0 \end{bmatrix} \in \mathbb{R}^{n \times p}$$

$$\sigma_0 > \sigma_1 > \dots > \sigma_{p-1} > 0$$

$$\tilde{y}_{OLS} = \left(\sum_{j=0}^{p-1} u_j u_j^T \right) y$$

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

$$\sigma_i^2 > 0$$

$$\frac{\partial^2 C}{\partial \beta \partial \beta'} = \frac{2}{2} X^T X$$

covariance matrix $\propto \cancel{X}^T \cancel{X}$

$\mathrm{var}(\beta_j)_{MSE} \propto (X^T X)^{-1}_{jj}$

$$\tilde{y}_{Ridge} = \left(\sum_{j=0}^{p-1} \frac{\sigma_j^2}{\sigma_j^2 + \lambda} u_j u_j^T \right) y$$

optimizing

$$C(\beta) = \frac{1}{n} \sum_{i=0}^{n-1} \left(y_i - \sum_{j=0}^{p-1} x_{ij} \beta_j \right)^2 + \lambda \sum_{j=0}^{p-1} \beta_j^2$$

$$\hat{\beta}_{Ridge} = (X^T X + \lambda \mathbb{I}_{p\times p})^{-1} X^T y$$

$$\lambda > 0$$

$$X = \mathbb{1}_{n\times n} \quad p = n$$

$$\tilde{y} = X\beta = \beta$$

OLS :

$$C(\beta) = \frac{1}{n}\sum_{i=0}^{n-1}(y_i - \beta_i)^2$$

$$\frac{\partial C}{\partial \beta_i} = 0 \Rightarrow \hat{\beta}_i = y_i = \tilde{y}_i$$

$$C_{Ridge} = \frac{1}{n} \sum_{i=0}^{n-1} (y_i - \beta_i)^2 + \lambda \sum_{i=0}^{n-1} \beta_i^2$$

$$\frac{\partial C_{Ridge}}{\partial \beta_i} = 0 \Rightarrow$$

$$\hat{\beta}_i^{Ridge} = \frac{y_i}{1+\lambda}$$

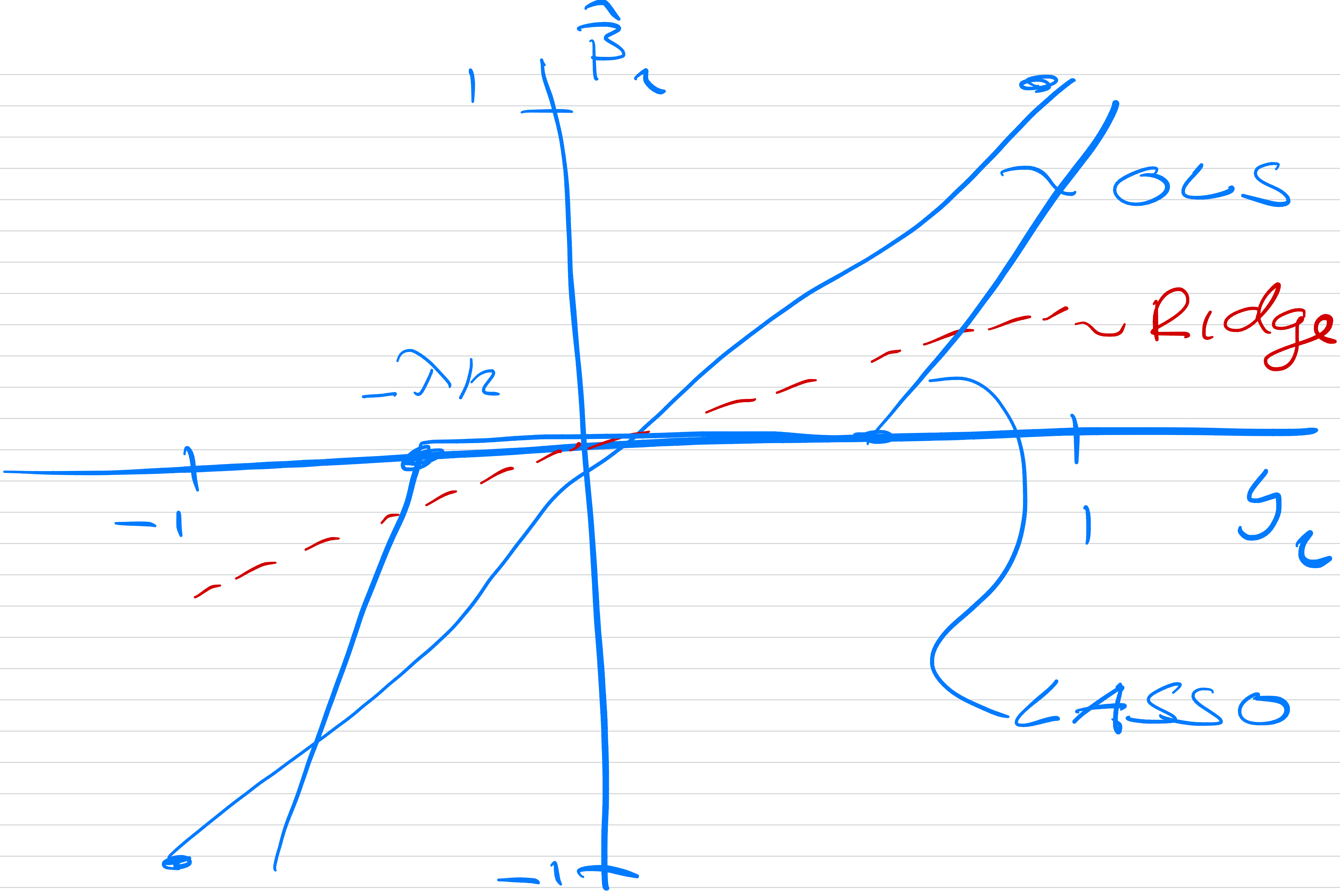
LASSO = Least absolute ~~shr~~shrinkage and selection operator

$$Q(\beta) = \frac{1}{n} \sum_{i=0}^{n-1} \left(y_i - \sum_{j=0}^{p-1} x_{ij} \beta_j \right)^2 + \lambda \underbrace{\sum_{j=0}^{p-1} |\beta_j|}_{\|\beta\|_1}$$

$$X = 1$$

$$\frac{\partial C}{\partial \beta_i} = -\frac{2}{n}\left(y_i - \beta_i\right) + \lambda \frac{\beta_i}{|\beta_i|} = 0$$

$$\hat{\beta}_i^{\,lasso} = \begin{cases} y_i - \lambda/2 & \text{if } y_i > \lambda/2 \\ y_i + \lambda/2 & \text{if } y_i < -\lambda/2 \\ 0 & \text{if } |y_i| \le \lambda/2 \end{cases}$$



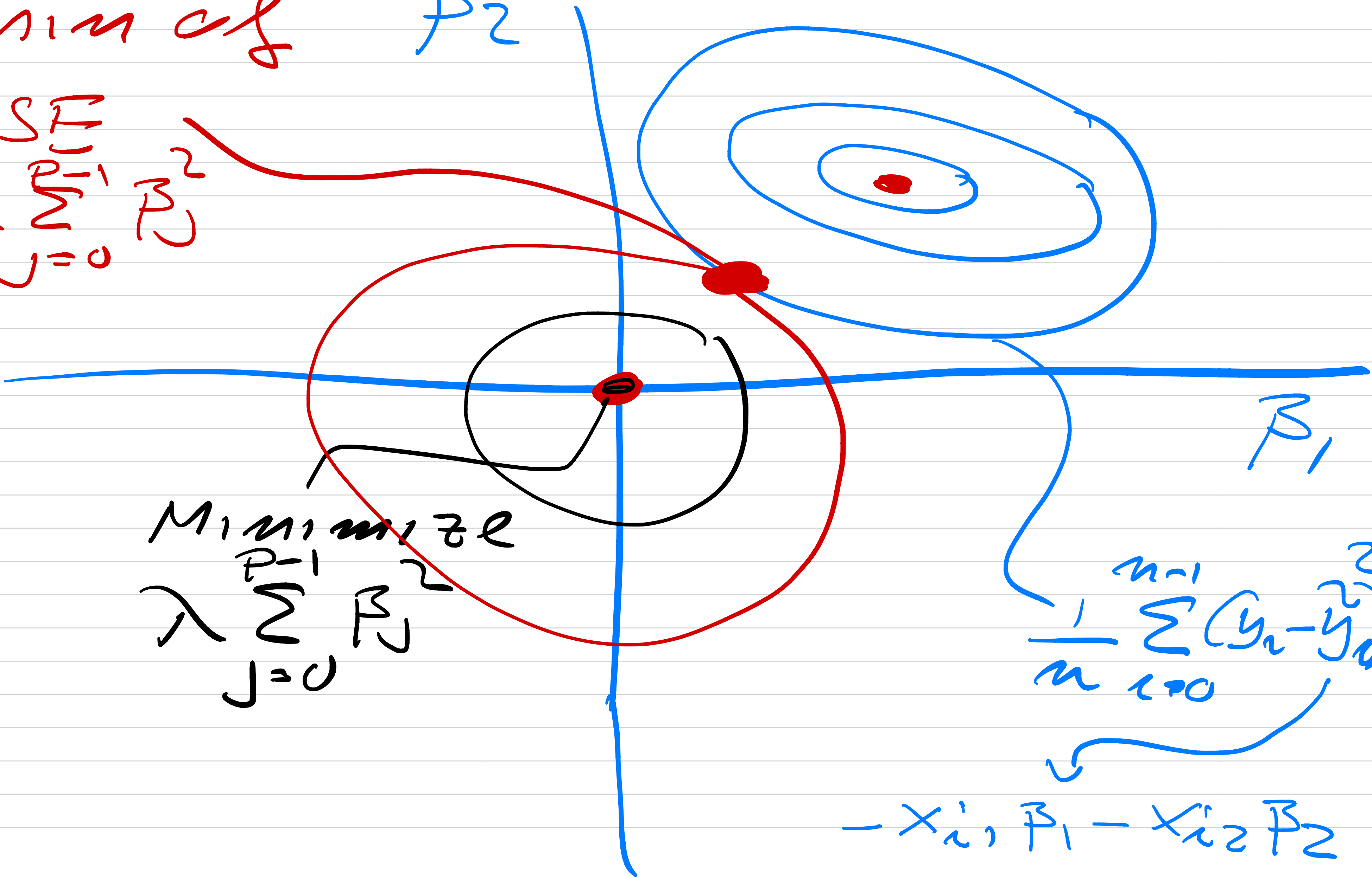
$\hat{\beta}_1$
1
$-\lambda/2$
-1
1
y_1
-1
OLS
Ridge
LASSO

(leave out β_0) β_1 β_2

min of

MSE

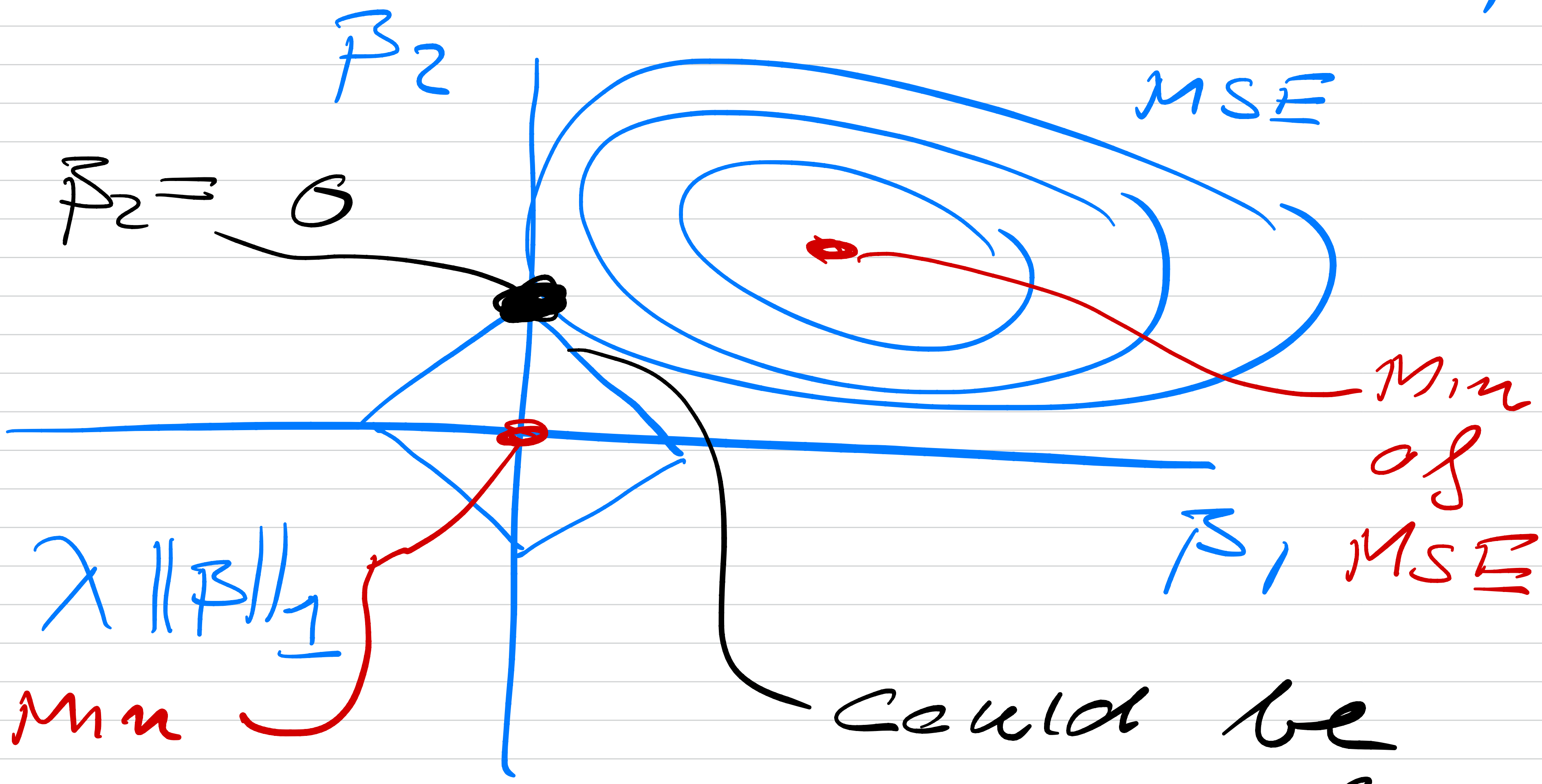
$+\lambda \sum_{j=0}^{p-1} \beta_j^2$



$$\frac{1}{n}\sum_{i=0}^{n-1}(y_i - \hat{y}_i)^2$$

$$- x_{i1}\beta_1 - x_{i2}\beta_2$$

Lasso (Raschka et al 122-125)



Could be min of $MSE + \lambda \|\beta\|_1$

Statistics

- Expectations, variance, covariance, PDF

Basic assumption for regression problems

$$y = \underbrace{f(x)}_{\text{stochastic function}} + \varepsilon \sim N(0, \sigma^2)$$

~~non-~~

Expectation values

$$\mathbb{E}[\varepsilon] = \mu_\varepsilon = \int_{x \in D} dx\, \varepsilon(x) P_\varepsilon(x)$$

mean value

$$= 0$$

$$\mathrm{var}[\varepsilon] = \int_{x \in D} (x - \varepsilon)^2 P_\varepsilon(x)\, dx$$

$$= \sigma^2$$

$$P_\varepsilon(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-x^2/2\sigma^2\right)$$

$$\tilde{y} = X\beta \cong f(x)$$

$$y \simeq \underbrace{X\beta}_{\text{non-stochastic}} + \varepsilon$$

$$y_i \simeq \underbrace{\sum_{j=0}^{p-1} x_{ij}\,\beta_j}_{x_{i*}\beta} + \varepsilon_i$$

$$\mathbb{E}[y_i] = \underbrace{\mathbb{E}[x_{i*}\beta]}_{x_{i*}\beta} + \underbrace{\mathbb{E}[\varepsilon_i]}_{=0}$$

$$\mathrm{var}[y_i] = \sigma_\varepsilon^2 \left(\begin{array}{c}\text{next}\\\text{week}\end{array}\right)$$

can infer

that

$$y_i \sim \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y - x_i * \beta)^2}{2\sigma^2}\right)$$