Week 8 Seminar - Regression

6 мая 2022 г. 19:06

$$y = ax + b$$

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} \qquad \bar{X} = N\begin{pmatrix} \begin{pmatrix} x_{11} \\ x_{12} \\ \vdots \\ x_{1N} \end{pmatrix} & \cdots & \begin{pmatrix} x_{k1} \\ x_{k2} \\ \vdots \\ x_{kN} \end{pmatrix} \end{pmatrix}$$

$\longrightarrow$ features

$$1D: \quad X = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_N & 1 \end{pmatrix} \qquad \begin{pmatrix} \theta_1 \\ \theta_0 \end{pmatrix} \begin{matrix} = a \\ = b \end{matrix}$$

$$\vec{Y} = \bar{X}\vec{\theta} + \vec{\varepsilon} \leftarrow \text{noise}$$

$$\begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

estimate $\Rightarrow$? $\hat{\theta}$

$$\boxed{\hat{Y} = \underline{X}\,\hat{\theta}}$$

$$\vec{\varepsilon} = \|Y - \hat{Y}\|_2^2 \to \min_{\theta}$$

$$\parallel$$

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

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

$$Y = \boxed{X}\theta + \varepsilon$$

$$\underset{k\times 1}{\hat{\theta}} = \underset{k\times n}{B}\,\underset{n\times 1}{Y} \quad ?$$

$$\underbrace{E(\hat{\theta}) = \theta}_{\text{undiased}}$$

$$\underset{\overset{\parallel}{\theta}}{E(\hat{\theta})} = E(B\cdot Y) = E\left(B\cdot(\underbrace{X\cdot\theta+\varepsilon}) \right)$$

$$E(B\cdot X\cdot\theta + B\cdot\varepsilon)$$

$$= B\cdot X\cdot\theta \quad B\cdot \underset{\overset{\parallel}{0}}{E(\varepsilon)}$$

$$\underset{k\times 1}{\theta} = \underbrace{B\cdot X}_{k\times k}\cdot\underset{k\times 1}{\theta}$$

identity matrix E I

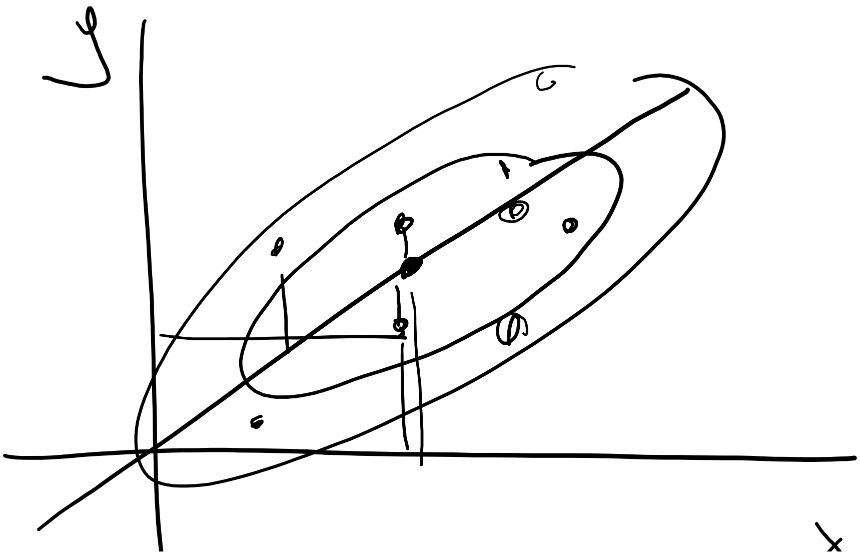
$$\underset{k\times n}{B}\cdot\underbrace{\underset{n\times k}{X}} = \underset{k\times k}{I}$$

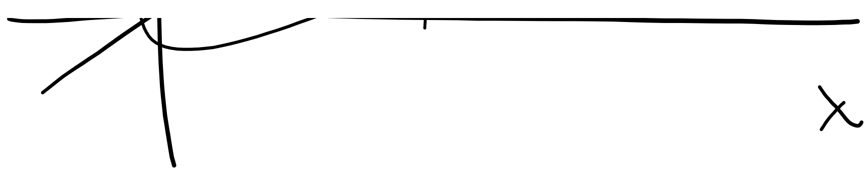
$$\boxed{\underset{k\times n}{B}\cdot\underset{n\times k}{X}=\underset{k\times k}{I}}$$

$$\boxed{\underset{n\times n}{A^{-1}}\cdot\underset{n\times n}{A}=\underset{n\times n}{I}}$$

pseudo - inverse of X matrix

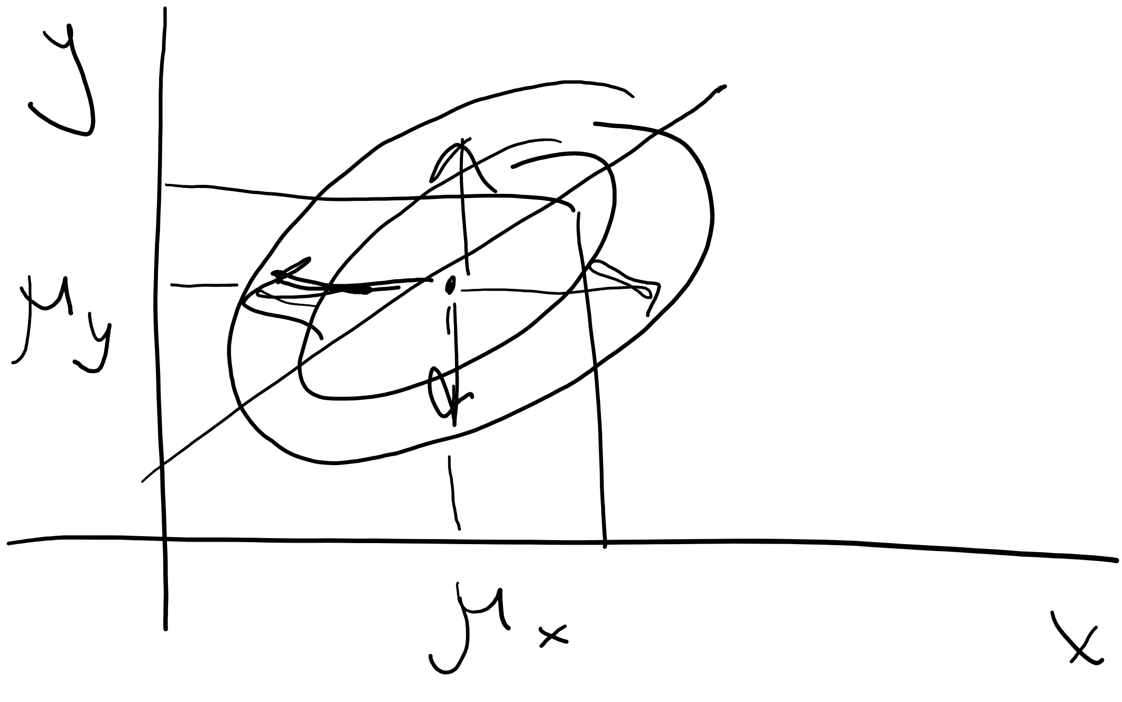
$$\begin{cases} Var(\hat{\theta}) \to \min \\ s.t.\quad BX = I \end{cases}$$





$$\mathbb{E}(y \mid x = \underline{x}) = r(x)$$

regression function



$$\mathcal{N}\left(\vec{\mu} = \begin{pmatrix}\mu_x \\ \mu_y\end{pmatrix}, \Sigma = \begin{pmatrix}\sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2\end{pmatrix}\right)$$

Pearson's correlation coefficient

$$\Sigma_{ij} = \mathbb{E}(x_i - \mathbb{E}(x_i))(x_j - \mathbb{E}(x_j))$$

$$\Sigma_{12} = \mathbb{E}\left((x_1 - \mathbb{E}x_1)(x_2 - \mathbb{E}x_2)\right)$$

$$\rho = \frac{\mathrm{Cov}(x_1, x_2)}{\sqrt{\mathrm{Var}(x_1)\mathrm{Var}(x_2)}} = \frac{}{\sigma_1 \sigma_2}$$

$$p(y \mid x) = \frac{p(y, x)}{p_{11}(x)}$$

$$\int_y p(y, x)\, dy$$

$$r(x) = \underline{\mu}_x + \underline{\rho}\,(x - \underline{\mu}_y)\frac{\underline{\sigma_x}}{\underline{\sigma_y}}$$

$$\operatorname{Var}(\hat{\theta}_i) =$$

$$= \ldots = b_i^T \, \mathbb{E}(\underbrace{\varepsilon \varepsilon^T}_{=}) b_i$$

$$i \begin{pmatrix} & j & \\ & \cdot & \\ & & \end{pmatrix} = i \begin{pmatrix} \circ \end{pmatrix} \begin{pmatrix} \circ & \\ b & \\ & \end{pmatrix}$$

$$\mathcal{L} = b_i^T b_i + \Lambda_i (\ldots)$$

$$(\, i \quad) \begin{pmatrix} \circ \end{pmatrix}$$

$$\left(\frac{\partial \mathcal{L}}{\partial b_i} \right)_j$$

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

$\left. x_n \right)$

$$\mathcal{L} = X^T X = \sum_i x_i^2$$

$$\frac{\partial \mathcal{L}}{\partial \vec{x}} = \left(\frac{\partial \mathcal{L}}{\partial x_i} \right)$$

$$\frac{\partial \mathcal{L}}{\partial x_i} = 2x_i$$

$$2 \left(x_i \right) = 2\vec{x}$$

$$Y = X\vec{\theta} + \varepsilon$$

$$\underset{k \times 1}{\hat{\theta}} = \underset{k \times n}{B} \underset{n \times 1}{Y}$$

$$\boxed{B = (X^T X)^{-1} X^T}$$

pseudoinverse of X

$-1 \qquad T$

$$\underbrace{\underbrace{\left(\underset{k\times n}{X^T}\;\underset{n\times k}{X}\right)^{-1}}_{k\times k}\;\underset{k\times n}{X^T}}_{k\times n}$$

$$X = \begin{pmatrix} x_{11} & \cdots \\ x_{12} & \cdots \\ \vdots & \\ x_{1n} & \cdots \end{pmatrix}$$

points ↓ observations

features, k →

typically, $k \le n$

$$\operatorname{rank}\left(X \right) \le k$$

if rank$(X) < k$

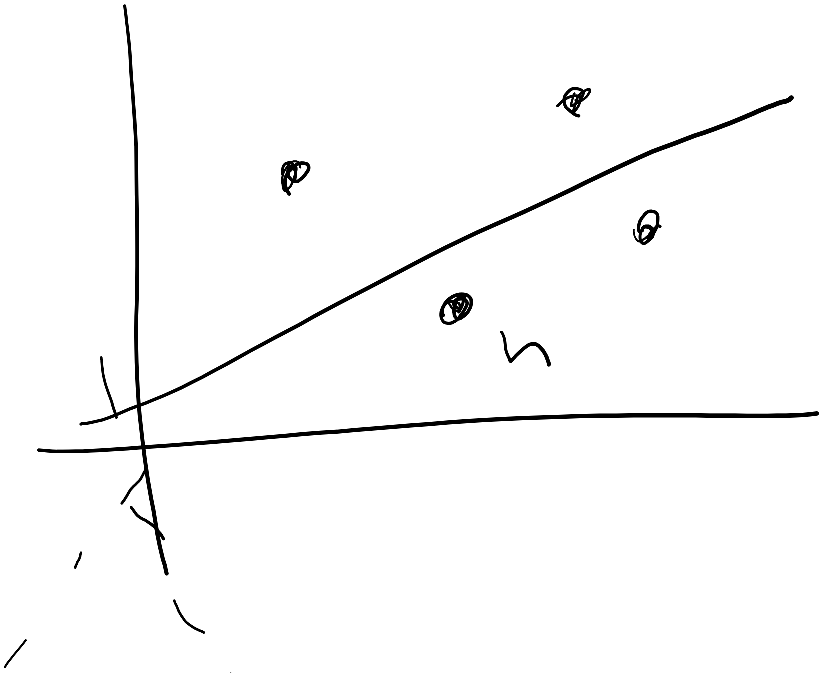
Ridge regression

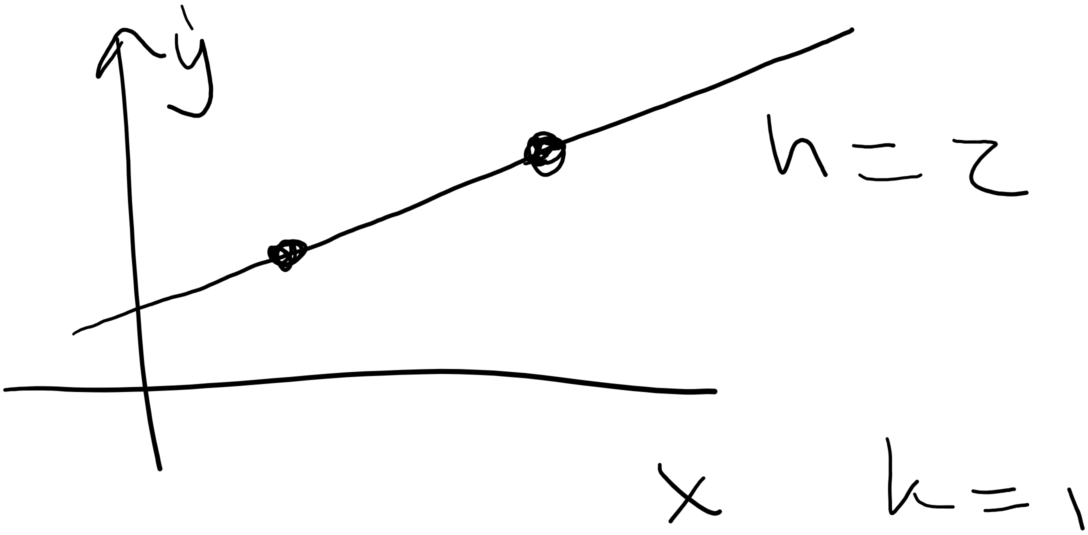
$$(\underbrace{X^T X}_{k \times k} + \lambda \underbrace{I}_{k \times k})^{-1} X^T$$

$$\hat{\theta} = B \cdot Y$$

$$Y = X \theta + \varepsilon$$

$$\hat{Y} = X \hat{\theta} = \underline{X} \cdot \underline{B} \cdot Y$$





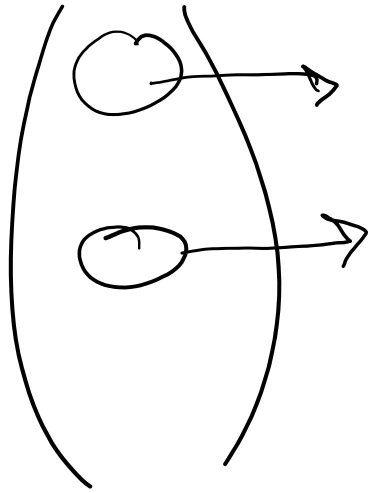
$$\frac{\overbrace{X_1}^{\sim N(0,\,\sigma^2)} + X_2 + \dots + X_n}{\sigma^2} \sim \chi^2(n)$$

find:

fit: $B = (X^T X)^{-1} X^T$

$\hat{\theta} = BY$

predict:

$Y =$ 

Y_{train} Y_{test}

$\hat{Y}_{test} = X_{test(train)} \hat{\theta}$

R^2-score

$R^2 = 1 - \frac{D(y|x)}{}$

$$D(y)$$

$$\underbrace{(X^T X)}^{-1}$$

$$\text{cond. number} = \frac{|\lambda_{max}|}{|\lambda_{min}|}$$