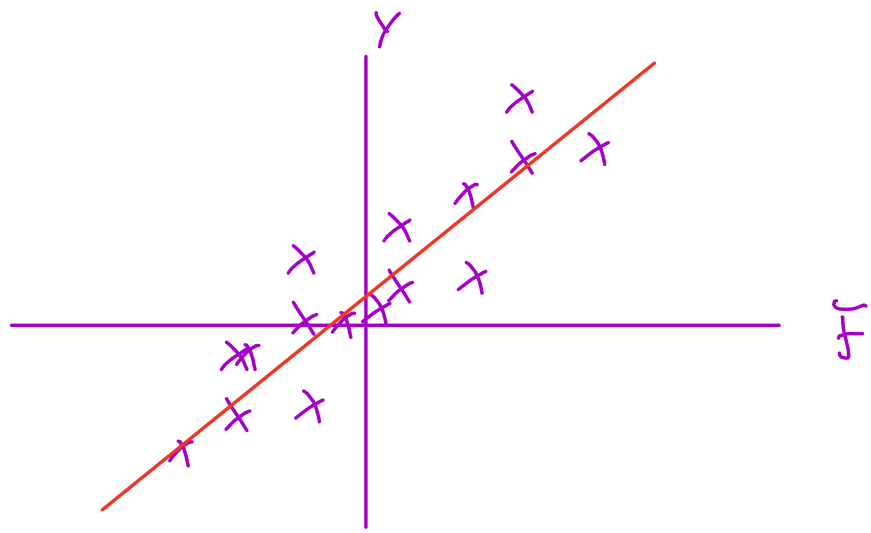
Question Sessions: PYS B-37

6:30

① Linear Regression

$T = \{(x_i, y_i)\}_{i=1}^{N}$ $\quad x_i \in \mathbb{R}^p$

feature space $\mathcal{F}$.



Goal: Find $f(x)$ such that

$$Y = f(x)$$

We typically parameterize the model such as

$$Y = f(x) = X\beta$$

unknown parameters

② Motivating Question:

Given a test point x_0 find

$(x_0, y_0) \sim p(x, y)$ the joint distribution.

③ Review of LM

$$Y = f(x) + \varepsilon \qquad f: \mathbb{R}^p \mapsto \mathbb{R}.$$

$$f(x) = X^T\beta$$

Best guess for f based on OLS

$$\boxed{\hat{\beta}_{OLS} = (X^TX)^{-1}X^TY}$$

Best guess for f wrt Ridge Regression

$$\boxed{\hat{\beta}_R = (X^T X + \lambda I)^{-1} X^T Y}$$

<u>Rmk:</u> If columns of X are orthonormal

then $\hat{\beta}_{Ridge} = \left(\frac{1}{1+\lambda}\right)\hat{\beta}_{OLS}$

Great for X's that are highly multicolinear.

Assume that the data is centered and we fit a no-intercept model.

<u>Rmk:</u>

$$\underset{n \times p}{X} = \underbrace{\underset{n \times p}{U}}_{\text{orthonormal columns}} \underset{p \times p}{\Sigma} \underbrace{\underset{p \times p}{V^T}}_{\text{orthonormal columns}}$$

$\text{colspan}(X) = \text{colspan}(U)$

Colspace(X) = Colspace(U)

Therefore

$$\hat{y}_{OLS} = U\Sigma V^T (V\Sigma^2 V^T)^{-1} V\Sigma U^T y$$

$$= U\Sigma V^T (V\Sigma^{-2} V^T) V\Sigma U^T y$$

$$= UU^T y$$

$$= \boxed{\sum_{j=1}^{p} u_j u_j^T y}$$

$$\hat{y}_R = X\hat{\beta}_R = U\Sigma V^T (V\Sigma^2 V^T + \lambda I)^{-1} V\Sigma U^T y$$

$$= U\Sigma V^T (V(\Sigma^2 + \lambda I)V^T)^{-1} V\Sigma U^T y$$

$$= U\Sigma (\Sigma^2 + \lambda I)^{-1} \Sigma U^T y$$

$$= \boxed{\sum_{j=1}^{p} \left(\frac{\sigma_i^2}{\sigma_i^2 + \lambda}\right) u_j u_j^T y}$$

Shrinkage in Colspace(X)