Homework Q's

4.2 $f(\vec{u},\vec{v}) = \underbrace{\vec{u}^T A \vec{u}} + \underbrace{\vec{v}^T B \vec{u}} + c$

use defns.

$$\vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \quad A = \begin{bmatrix} A_{11} & A_{12} & \\ A_{21} & A_{22} & \\ & & \ddots \end{bmatrix}$$

12.3

$$\frac{1}{n-1} \sum_{i=1}^{n} \underbrace{(\vec{x}_i - \vec{\mu})(\vec{x}_i - \vec{\mu})^T}_{\text{outer product } = 15 \times 15} \qquad \vec{x}_i \quad 15\text{-dim}$$

$$\begin{bmatrix} | & & | \\ \vec{\mu} & \cdots & \vec{\mu} \\ | & & | \end{bmatrix}$$

broadcasting

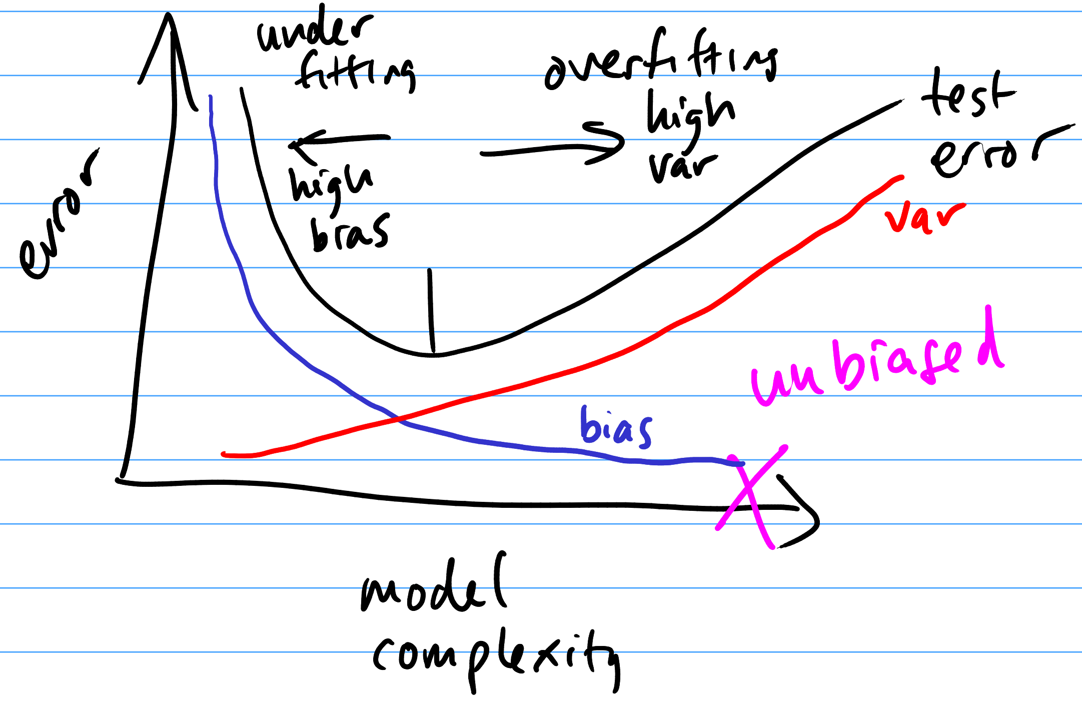
A = np.array(...)

X - mu[:, np.newaxis]

np.repeat

A.shape (100,) (100, 1)

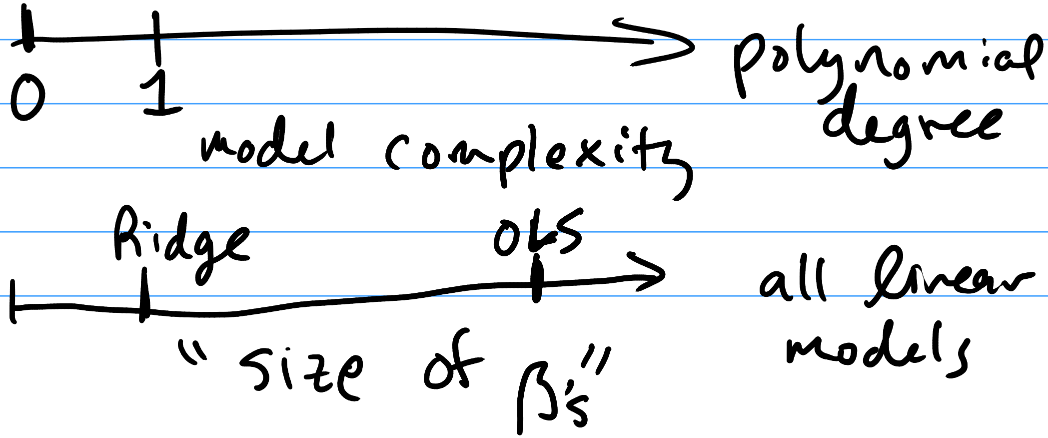
Problem 11 → new PDF



big coefficients on polynomials

β's $\gg 100$

Ridge Regression



Ordinary least squares (OLS) is <u>unbiased</u>

Assume: X^TX not singular

$$\vec{y} = X\vec{\beta}^* + \vec{\varepsilon}$$

true model

noise

np.random.randn

Gaussian random vars

$\mathbb{E}[\vec{\varepsilon}] = 0$

$$\vec{\beta} = (X^TX)^{-1}X^T\vec{y} = (X^TX)^{-1}X^T(X\vec{\beta}^* + \vec{\varepsilon})$$

$$= (X^TX)^{-1}X^TX\vec{\beta}^* + (X^TX)^{-1}X^T\vec{\varepsilon}$$

$$= \vec{\beta}^* + (X^TX)^{-1}X^T\vec{\varepsilon}$$

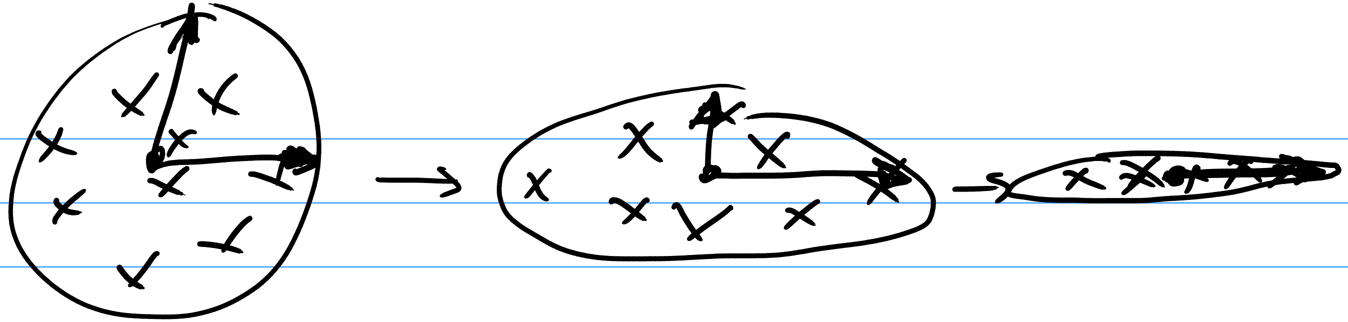
$$\boxed{\mathbb{E}[\vec{\beta}] = \vec{\beta}^*} + (X^TX)^{-1}\cancelto{0}{X^T\mathbb{E}[\vec{\varepsilon}]}$$

averaging over noise, OLS estimate = truth

$$\text{Bias} = \mathbb{E}[\vec{\beta}] - \vec{\beta}^*$$

Observations

(Cholesky decomposition generating correlated Gaussians)



$\sigma_1 \approx \sigma_2$

$\sigma_1 \approx 5\sigma_2$ same rough size

$\sigma_1 >> \sigma_2$, $\sigma_1 \approx 100\sigma_2$, small σ_2

has noise

$$\vec{\beta}_{OLS} = \sum_{i=1}^{\mathrm{rank}(X)} \vec{v}_i \underbrace{\left(\frac{\vec{u}_i^{\,T} \vec{y}}{\sigma_i} \right)}_{w_i}$$

$\frac{1}{\text{small}} = \text{big}$

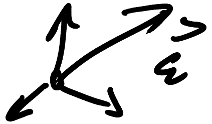
noise : equally shared among the components

Correlations in X $\Longleftrightarrow$ small s.v.'s

If linear model is true :

$$w_i = \underbrace{\vec{v}_i^T \vec{\beta}^*}_{i^{th} \text{ component } \vec{\beta}^* \text{ in } V \text{ basis}} + \underbrace{\frac{\vec{u}_i^T \vec{\varepsilon}}{\sigma_i}}_{\text{effect of noise}}$$

iid.

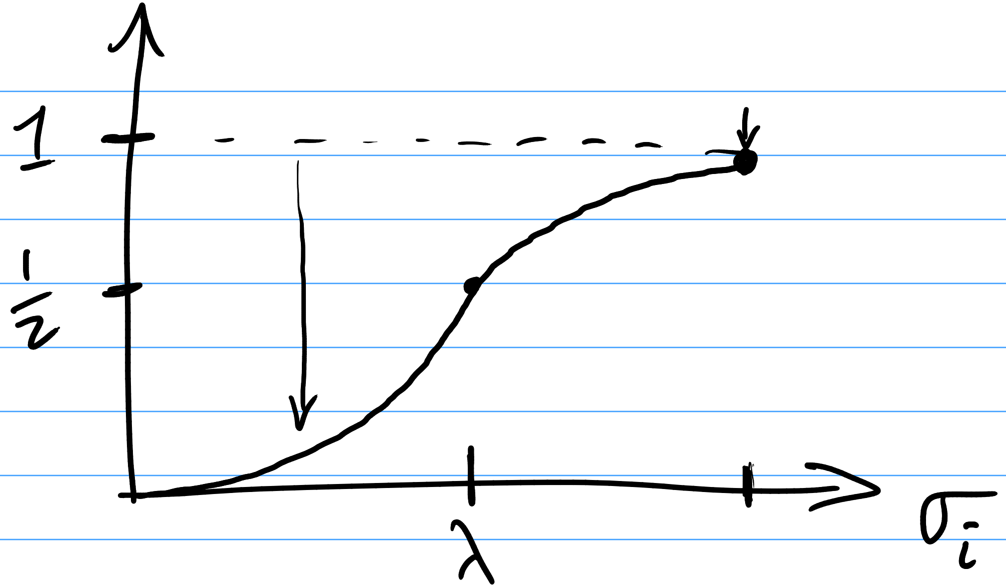


$$\vec{\beta}_{ridge} = \sum_{i=1}^{rank(X)} \vec{v}_i \underbrace{\left(\frac{\vec{u}_i^T \vec{y}}{\sigma_i + \lambda} \right)}_{w_i} \quad , \quad w_i = \underbrace{\left(\frac{\sigma_i}{\sigma_i + \lambda} \right)}_{\text{plot}} \vec{v}_i^T \vec{\beta}^* + \frac{\vec{u}_i^T \vec{\varepsilon}}{\sigma_i + \lambda}$$

λ ridge parameter, must be tuned for dataset / scenario

$\sigma_i > \lambda$: not much effect

$\sigma_i < \lambda$: have strong effect on those components



$$\frac{\sigma_i}{\sigma_i+\lambda} = \frac{1}{1+\frac{\lambda}{\sigma_i}}$$

Ridge : add bias ($\lambda = 0$ no bias, OLS)
reduce variance

Extra: How to get w_i formula

$$w_i = \frac{\vec{u}_i^T \vec{y}}{\sigma_i} = \frac{\vec{u}_i^T (X\vec{\beta}^* + \vec{\varepsilon})}{\sigma_i}$$ using linear assumption

$$= \frac{\vec{u}_i^T (USV^T \vec{\beta}^* + \vec{\varepsilon})}{\sigma_i}$$ using $X = USV^T$

$$= \underbrace{\vec{u}_i^T USV^T}_{\vec{u}_i^T U = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \end{pmatrix}} \vec{\beta}^* \left(\frac{1}{\sigma_i}\right) + \frac{\vec{u}_i^T \vec{\varepsilon}}{\sigma_i} = \underbrace{\vec{v}_i^T \vec{\beta}^*}_{i^{th} \text{ component in } V \text{ basis}} + \overbrace{\frac{\vec{u}_i^T \vec{\varepsilon}}{\sigma_i}}^{\text{noise term}}$$

i^{th} position

So it selects i^{th} singular vector, i.e.

$$\vec{u}_i^T USV^T \vec{\beta}^* = \sigma_i \vec{v}_i^T \vec{\beta}^*$$