FYS-STK3155/4155, Lecture September 15, 2025

FYS-STK3155/4155 September 15

Statistical interpretations
(Mainly OLS)
- Expectation values
- central limit theorem
- Resampling methods and
  Bias-variance tradeoff.

Basic assumptions

$$\underbrace{y}_{?} = \underbrace{f(x)}_{\text{non-stochastic functions}} + \underbrace{\varepsilon}_{N(0,\sigma^2)}$$

non-stochastic functions

?

$N(0, \sigma^2)$

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

stochastic variable x

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

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

$$\mu_x = \mathbb{E}[x] = \int_{x\in D} dx\, x\, P_x(x)$$

sample mean ($P_x(x)$ is unknown)

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

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

$$= \int_{x \in D} P_x(x)\, dx\, (x - \mu_x)^2$$

sample variance

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

$$\neq \mathrm{var}[x]$$

Design matrix

$$X = [\vec{x}_0 \ \vec{x}_1 \ \cdots \ \vec{x}_{p-1}]$$

$$\mathrm{cov}(\vec{x}_i, \vec{x}_j) =$$

$$\frac{1}{n} \sum_{\ell=0}^{n} \underbrace{(x_{\ell j} - \bar{\mu}_{x_j})(x_{\ell i} - \vec{\mu}_{x_i})}_{\text{number}}$$

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

SVD

$$X^T X v_i' = \sigma_i^2 v_i'$$

$$V = [\vec{v}_0 \ \vec{v}_1 \ \cdots \ \vec{v}_{p-1}]$$

$$\tilde{y} = \underbrace{X \cdot \theta}_{\text{non-stochastic}} \simeq f(x)$$

$$y_i = \underbrace{\sum_{j=0}^{p-1} x_{ij}\theta_j}_{X_{i*}\theta} + \varepsilon_i$$

$$\mathbb{E}[y_i] = ?$$

$$= \mathbb{E}[x_{i*}\theta] + \mathbb{E}[\varepsilon_i]$$

$$\approx 0$$

what does $P_y(y)$ look like?

$$\mathbb{E}[x_{i*}\theta] = \int_{y \in D} P_y(y)\, x_{i*}\theta \times dy$$

$$\int_{y \in D} P_y(y)\, dy = 1$$

$$\mathbb{E}[y_i] = X_{i*}\theta$$

$$= \sum_{j=0}^{P-1} x_{ij}\theta_j$$

$$\mathbb{E}[y] = X\theta$$

Assumption

$$y_i \sim \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_i - X_{i*}\theta)^2}{2\sigma^2}\right]$$

$$\sim N(X_{i*}\theta, \sigma^2)$$

$var[y_i] = \sigma^2$

Can we derive OLS eq using

$$P(y_i x \mid \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_i - x_i * \theta)^2}{2\sigma^2}\right]$$

$$= P_i$$

$$P(\vec{y}\,X \mid \theta) = \prod_{i=0}^{n-1} p_i$$

$y_i \sim$ i.i.d.

$$\hat{\theta} = \underset{\theta \in \mathbb{R}^p}{\operatorname{argmax}}\; P(\vec{y}\,X \mid \theta)$$

$$\vec{\nabla}_\theta\, P(\vec{y}\,X \mid \theta) = 0$$

$$C(\theta) = -\log P(\vec{y}\, X \mid \theta)$$

$$= - \sum_{i=0}^{n-1} \log (y_i X \mid \theta)$$

$$= + \frac{n}{2} \log 2\pi\sigma^2 + \sum_{i=0}^{n-1} \frac{(y_i - X_i * \theta)^2}{2\sigma^2}$$

$$\nabla_\theta C(\theta) = 0$$

$$C(\theta) = \frac{n}{2} \log 2\pi\sigma^2 + \frac{1}{2\sigma^2} \| y - X\theta \|_2^2$$

$$\nabla_\theta C(\theta) = 0 \implies$$

$$\hat{\theta} = (X^T X)^{-1} X^T y$$

Resampling with Bootstrap

Domain

$$D = [z_0 \; z_1 \; \dots \; z_{n-1}]$$

$$\bar{\mu}_z = \frac{1}{n}\sum_{i=0}^{n-1} z_i \quad (\neq \mu_z)$$

$$\mathrm{var}[z] \neq \frac{1}{n}\sum_{i=0}^{n-1}\left(z_i - \bar{\mu}_z\right)^2 = \bar{\sigma}_z^2$$

with replacement

$= z_0$ $\qquad\qquad$ z_0

Reshuffle data randomly

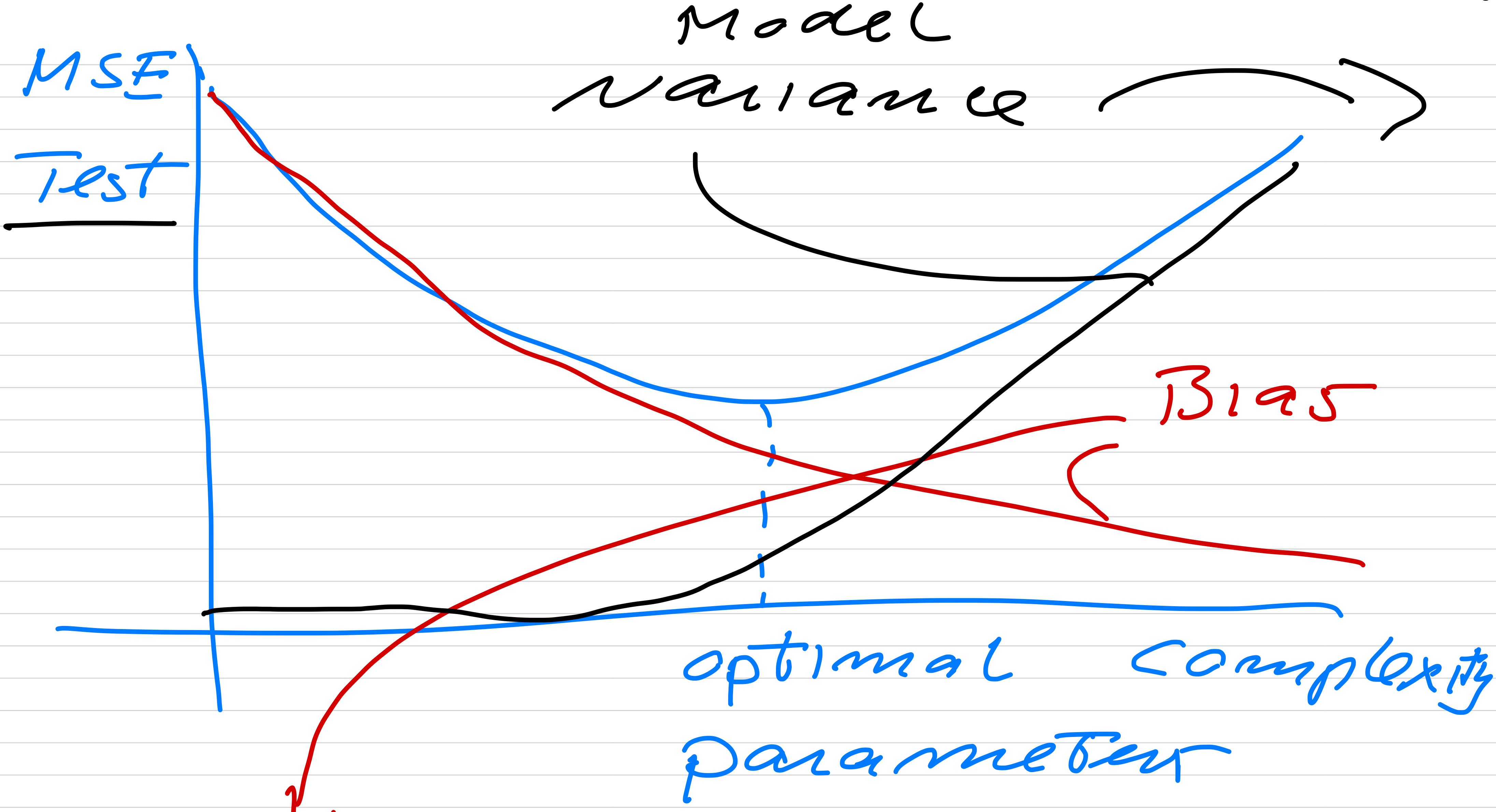
$$D' = [z_0' \; z_1' \; \dots \; z_i' \; \dots \; z_{n-1}']$$

Repeat B-times

$$\bar{\mu}_B = \frac{1}{B} \sum_{b=1}^{B} \bar{\mu}_b$$

hope is that this gets closer to true mean

$$\bar{\sigma}_B^2 = \frac{1}{B} \sum_{b=1}^{B} \left(\bar{\mu}_b - \bar{\mu}_B \right)^2$$

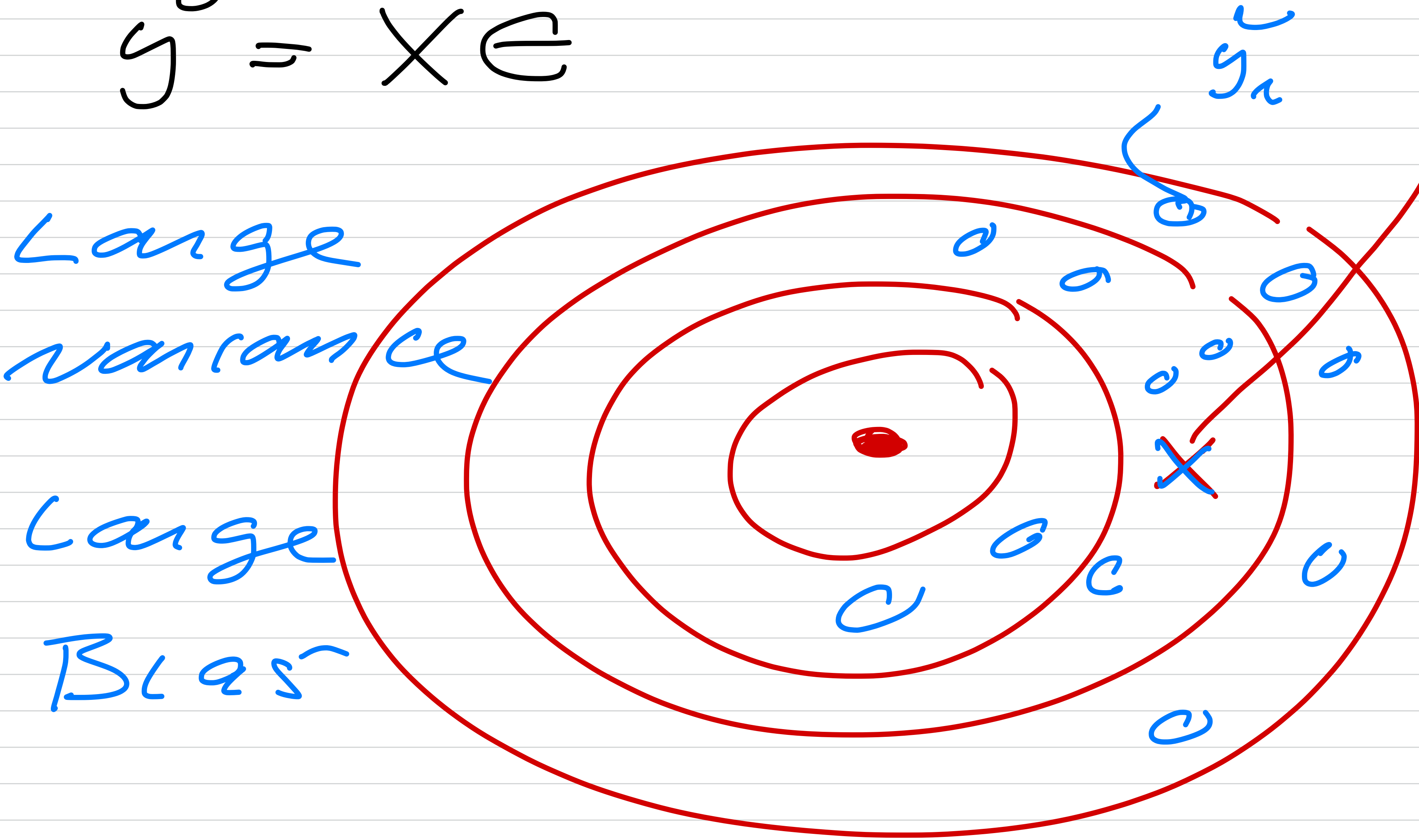


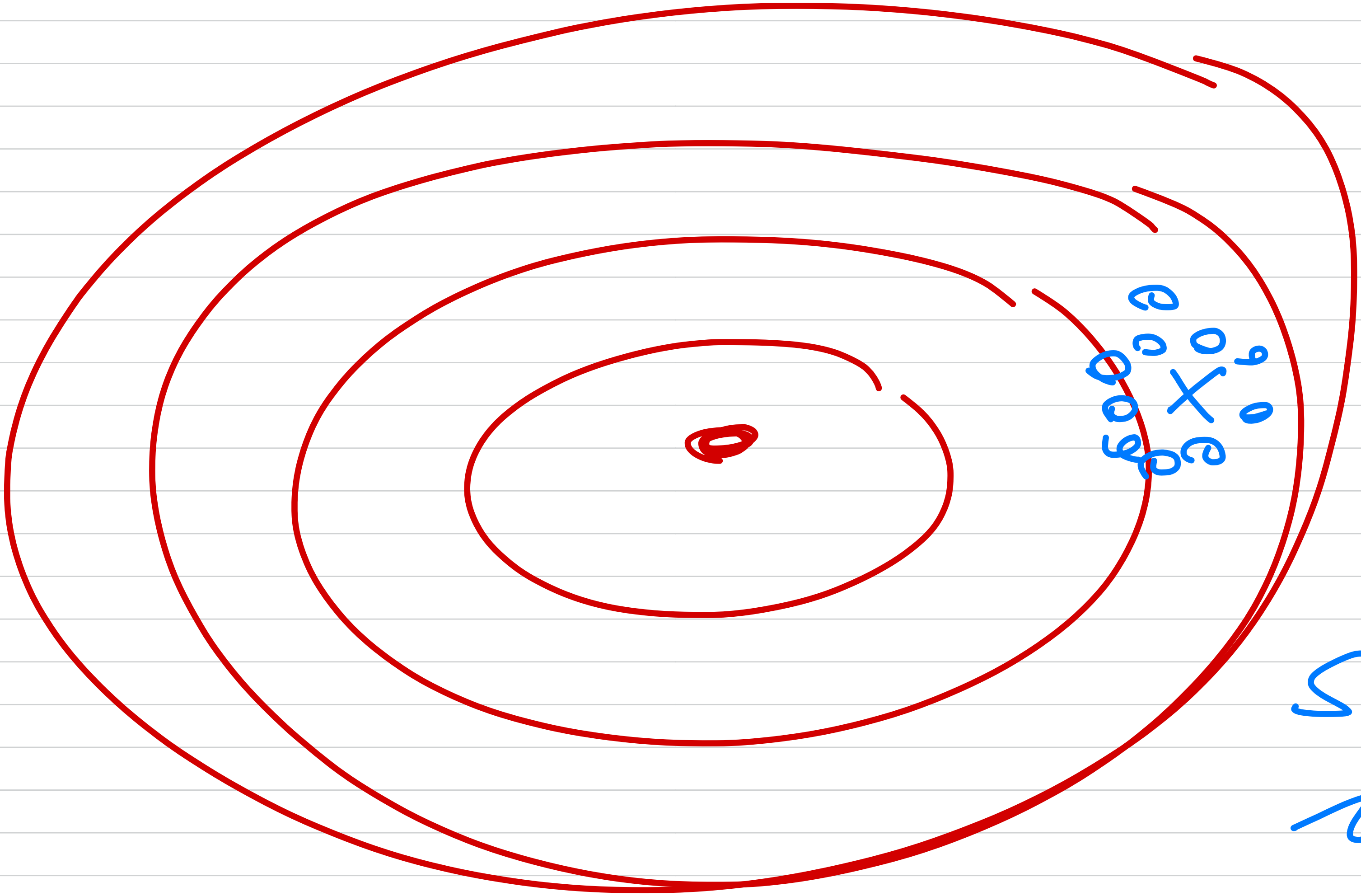
$$\frac{1}{n}\sum_{i=0}^{n-1}\left(y_i - \mathbb{E}[\tilde{y}]\right)^2$$

(f_i)

$$\mathrm{var}[\hat{y}] = \frac{1}{n}\sum_{i=0}^{n-1}\left(\hat{y}_i - \mathbb{E}[\hat{y}]\right)^2$$

$$\hat{y} = X\hat{\theta}$$

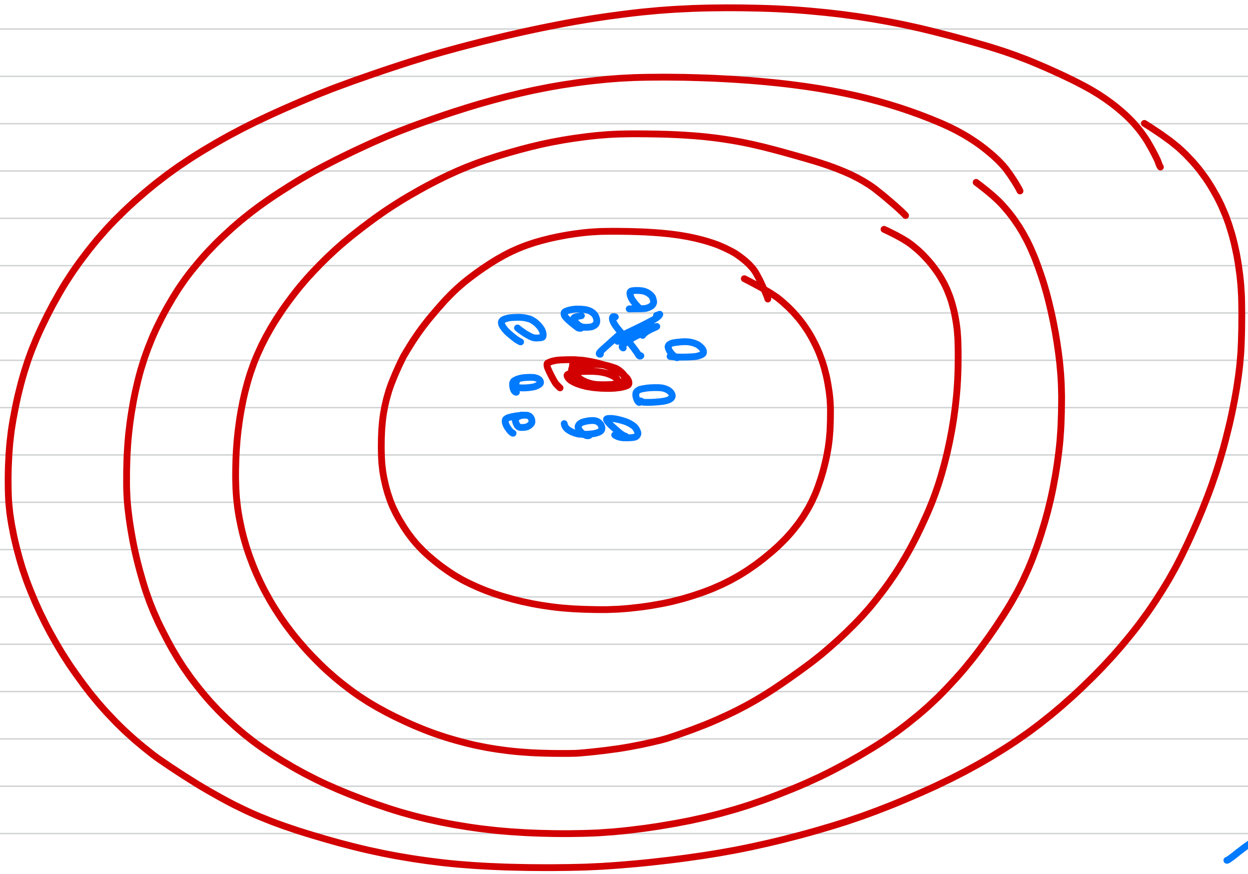




Small variance

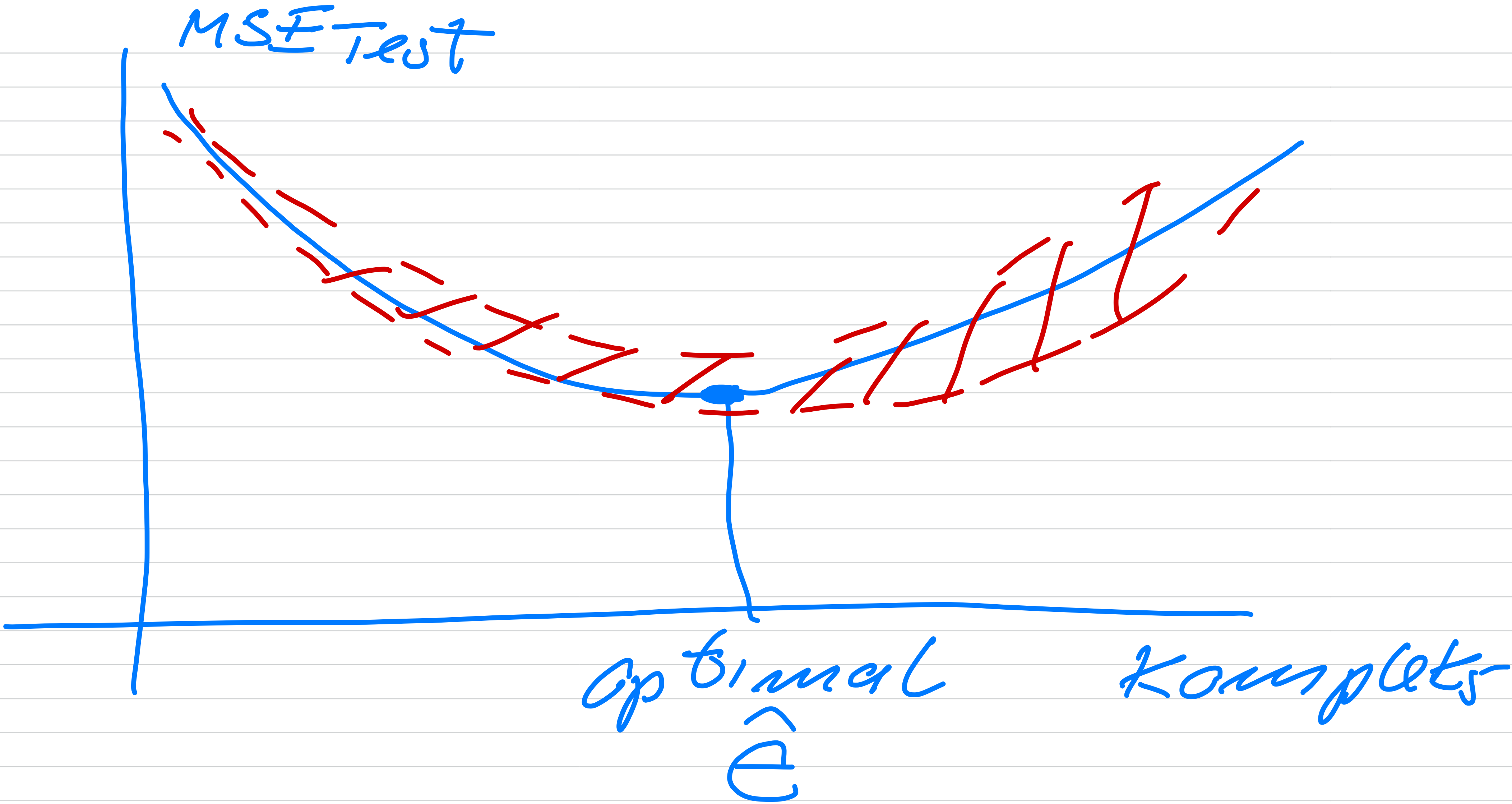
Large Bias

Ideal



Low
Variance

Low
Bias



MSE Test
optimal
ê
Komplex

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2 \qquad \tilde{y} = Xe$$

$$= \mathbb{E}[(y-\tilde{y})^2]$$

$$(\langle (y-\tilde{y})^2 \rangle)$$

$$= \mathbb{E}[y^2] - 2\,\mathbb{E}[y\tilde{y}]$$

$$+ \underbrace{\mathbb{E}[\tilde{y}^2]}$$

$$\int_{y\in D} y^2 p_y(y)\,dy$$

$$\mathbb{E}[y^2] = var[y] + (\mathbb{E}[y])^2$$

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

$$\mathbb{E}[y^2] = \mathbb{E}[(f+\varepsilon)^2]$$

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

$$\mathbb{E}[f^2] + 2\mathbb{E}[f \cdot \varepsilon] + \underbrace{\mathbb{E}[\varepsilon^2]}_{\sigma^2}$$

$$\mathbb{E}[f(x)] = \int f(x)\, P_y(y)\, dy$$

$$= f(x) \underbrace{\int P_y(y)\, dy}_{=1}$$

$$= f(x) \cdot 1$$

$$\mathbb{E}[f \cdot \varepsilon] = f \underbrace{\int \varepsilon \, d\varepsilon \, P_\varepsilon(\varepsilon)}_{=0}$$

$$= f \cdot \underbrace{\mathbb{E}[\varepsilon]}_{=0} = 0$$

$$\mathbb{E}[y^2] = f^2 + \sigma^2$$

$$\mathbb{E}[y\hat{y}] = \mathbb{E}[(f+\varepsilon)\hat{y}]$$

$$= \underbrace{\mathbb{E}[f\cdot\hat{y}]}_{f\,\mathbb{E}[\hat{y}]} + \underbrace{\mathbb{E}[\varepsilon\cdot\hat{y}]}_{\mathbb{E}[\varepsilon]\,\mathbb{E}[\hat{y}] \approx 0}$$

$$p(x,y) = p_1(x)\,p_2(y)$$

$$\mathbb{E}[xy] = \int dx\,dy\; p_1(x)\,p_2(y)\,xy$$

$$= \underbrace{\int dx\, p_1(x)\, x}_{\mathbb{E}[x]} \cdot \underbrace{\int p_2(y)\, y\, dy}_{\mathbb{E}[y]}$$

$$\mathbb{E}[(y-\hat{y})^2] =$$

$$f^2 - 2f\,\mathbb{E}[\hat{y}] + (\mathbb{E}[\hat{y}])^2$$

$$+ \ var[\hat{y}] + \sigma^2$$

$$\underbrace{\mathbb{E}[(f-\mathbb{E}[\hat{y}])^2]}_{\text{Bias}}$$

$$+ \ var[\hat{y}] + \sigma^2$$