Lecture Erasmus+ course on Machine Learning, October 23, 2023

Cost function

$$C(\beta) = \sum_{i=0}^{n-1} \mathcal{L}(y_i, x_i; \beta)$$

$$\left(\frac{1}{n} \sum_{i=0}^{n-1} (y_i - \tilde{y}_i)^2 \right.$$

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

$$MSE = \mathbb{E}\left(\|y - X\beta\|_2^2 \right) \leftarrow \text{sample MSE}$$

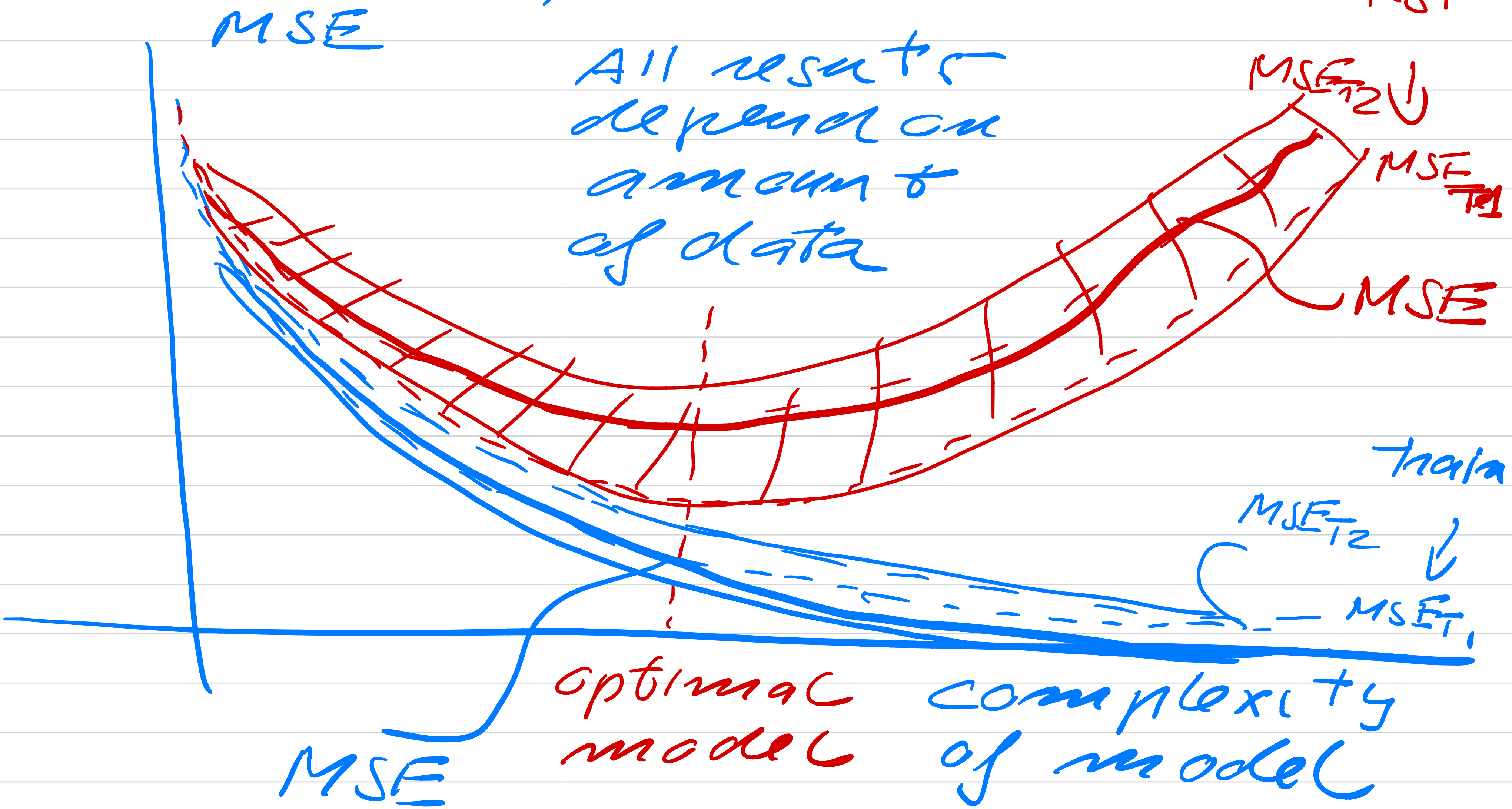
$$\mathbb{E}[x^n] = \int_{x \in D} x^n p(x) dx$$

unknown

$$\mu = \mathbb{E}[x] = \int_{x \in D} x p(x) dx$$

$$\neq \mu_x = \frac{1}{n} \sum_{i=0}^{n-1} x_i \quad \text{sample mean}$$

Calculate training MSE and Test MSE with different (but limited)



$$MSE_{Train} = \frac{1}{B} \sum_{i=1}^{B} MSE_{Train}(i)$$

$$\boxed{MSE_{Test} = \frac{1}{B} \sum_{i=1}^{B} MSE_{Test}(i)}$$

Resampling techniques are used to generate "new" data in order to obtain better estimates of statistical expectation values

Bootstrap (often small data sets $n \sim 100-1000$)

$$D = \{x_0, x_1, x_2, \ldots x_{n-1}\}$$

algorithm

(i) Reshuffle D randomly with replacement

$$D_1 = \{x_0^*, x_1^*, x_2^* \ldots x_{n-1}^*\}$$

$$D = \{1, 2, 3, 4\}$$

$$D_1 = \{2, 3, 1, 1\}$$

calculate expectation values, μ_i, σ_i^2 etc

(ii) Repeat (i) B-times

(iii) calculate

$$\mu = \frac{1}{B} \sum_{i=1}^{B} \mu_i$$

$$MSE = \frac{1}{B} \sum_{i=1}^{B} MSE(i)$$

Example: Gaussian with $n = 10^4$, $\mu = 100$, $\sigma = 15$

perform m- bootstrap

$\sigma_m \sim \sigma / \sqrt{m}$, follows

from central limit theorem.

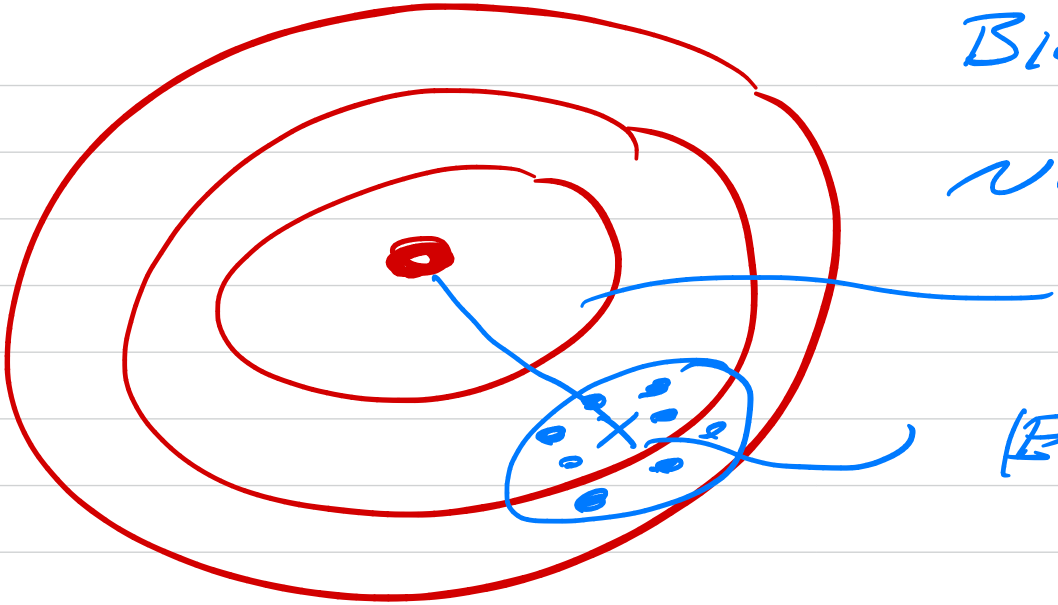
$\mu_n = \mu_m$

$$MSE = Bias + var[\tilde{y}] + \sigma^2$$

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

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

variance of model



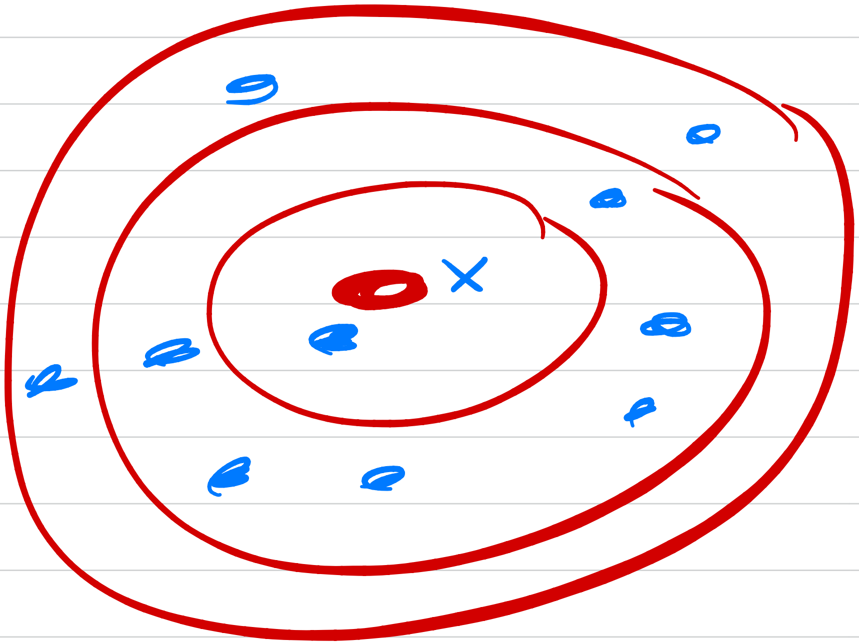
Bias high

variance low

large deviation

$\mathbb{E}[\hat{y}]$ from y

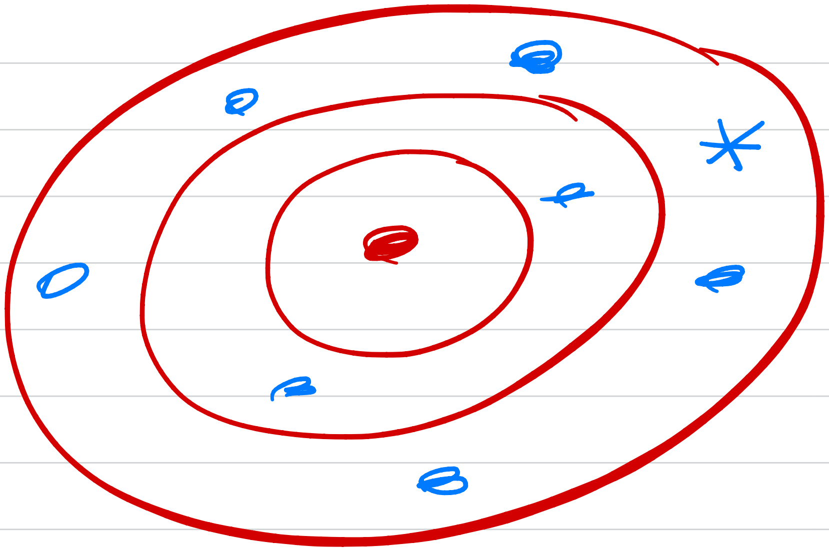
$\Rightarrow$ large Bias



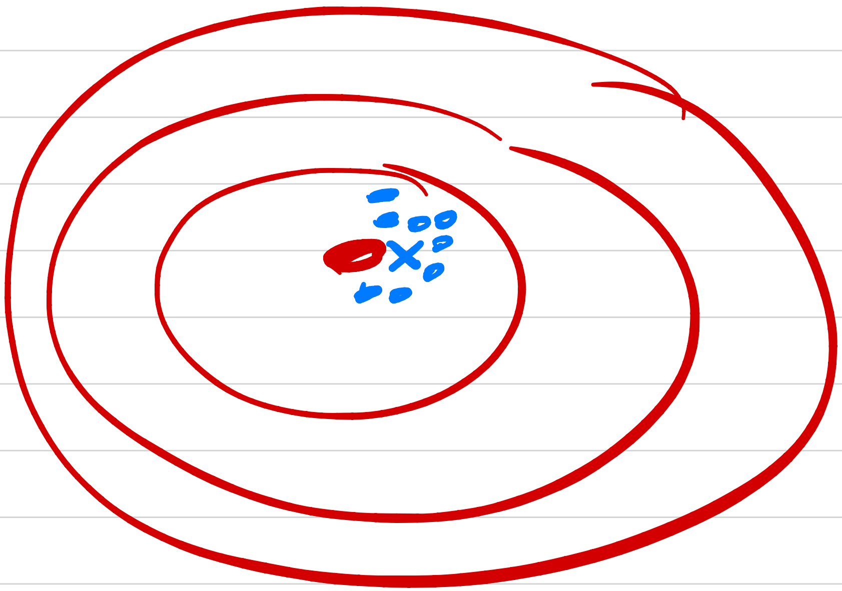
Low Bias

high variance

(overfitting)



high Bias
high variance
overfitting



Low Bias
Low variance
ideal