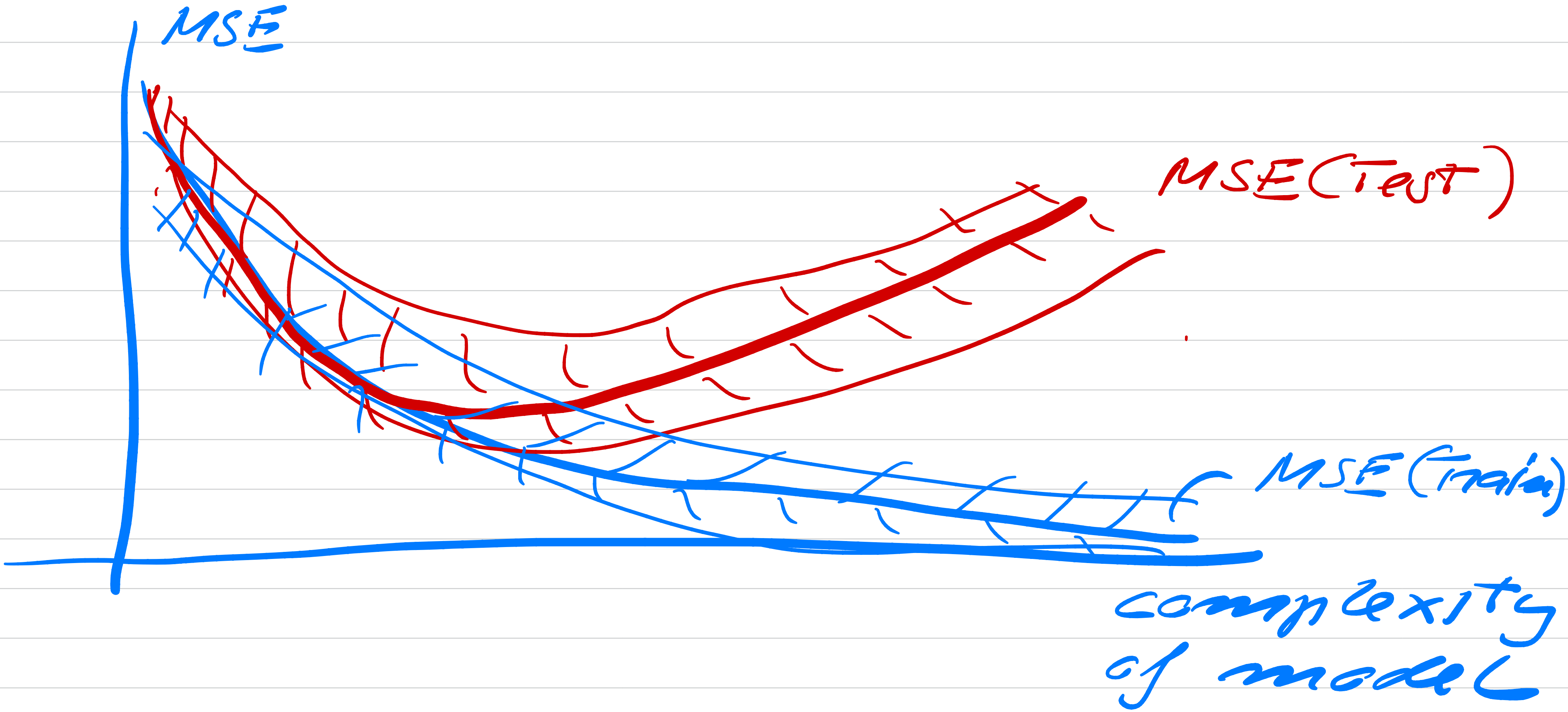
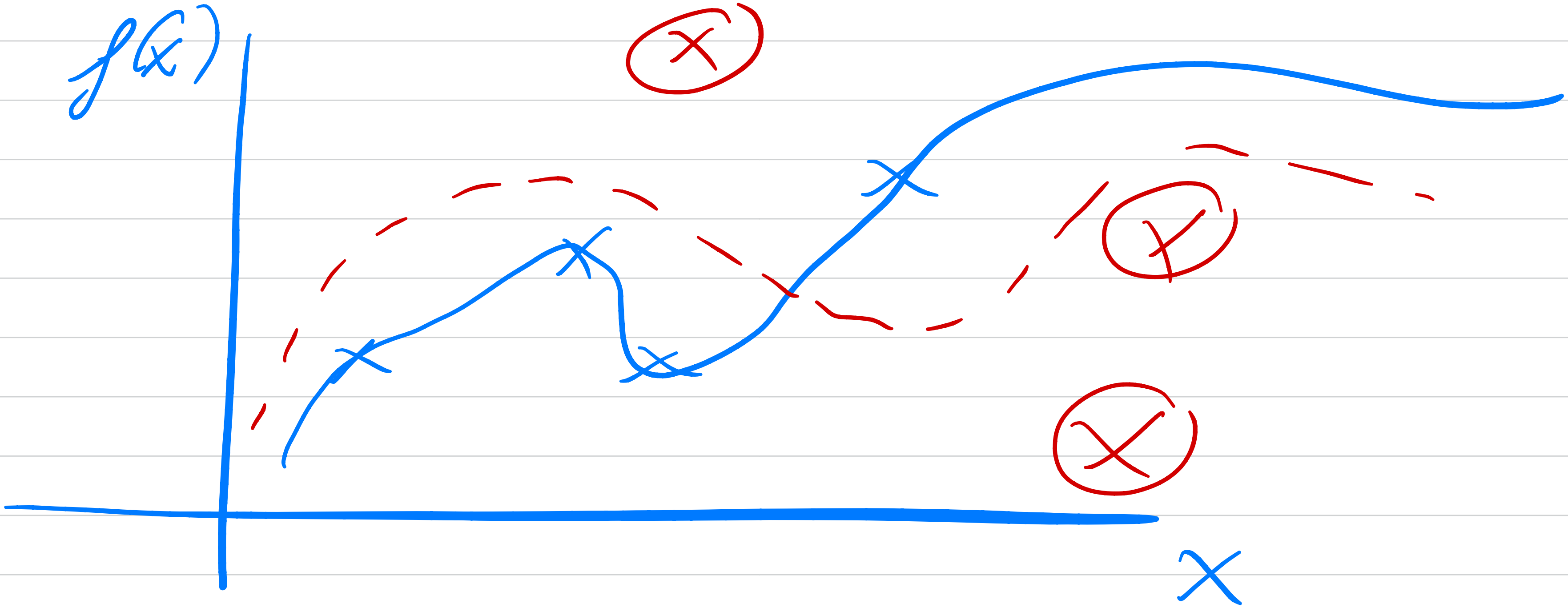
Erasmus+ course on Machine Learning, lecture October 30, 2023

Function fitting, MSE as a test of the model $\tilde{y}$

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





Resampling

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

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

Bootstrap method

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

FOR $i = 1, B$

reshuffle D

$D_i = \{x_0^*, x_1^*, -- x_{n-1}^*\}$

pick randomly n entries with replacement

calculate various expectation values

END

calculate final averages

Bias-variance tradeoff

$$MSE = \int p(x)\,(y-\tilde{y})^2\,dx$$

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

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

$$\left(\sum_{\substack{i=1\\ x_i\in D}}^{N} p(x_i)\,y^n(x_i)\right)$$

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

$$y(x) = f(x) + \varepsilon$$

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

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

nonstochastic function

$$\mathbb{E}[f^n] = \int_{z \in D} p(z)\, f(x)\, dz$$

$$= f(x) \underbrace{\int p(z)\, dz}_{=1}$$

$$= f^2 + 2f \underbrace{\mathbb{E}[\varepsilon]}_{=0} + \underbrace{\mathbb{E}[\varepsilon^2]}_{\sigma^2}$$

$$= f^2 + \sigma^2$$

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

$$= \mathbb{E}[f\tilde{y}] + \mathbb{E}[\varepsilon\tilde{y}]$$

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

$\tilde{y}$ and ε are independent of each other

Bringing everything together

$$f^2 - 2\mathbb{E} f\,\mathbb{E}[\tilde{y}] + \sigma^2 + \mathrm{var}[\tilde{y}]$$
$$+ (\mathbb{E}[\tilde{y}])^2$$

$$= \mathbb{E}((f - \mathbb{E}[\tilde{y}])^2) + \mathrm{var}[\tilde{y}]$$
$$+ \sigma^2$$

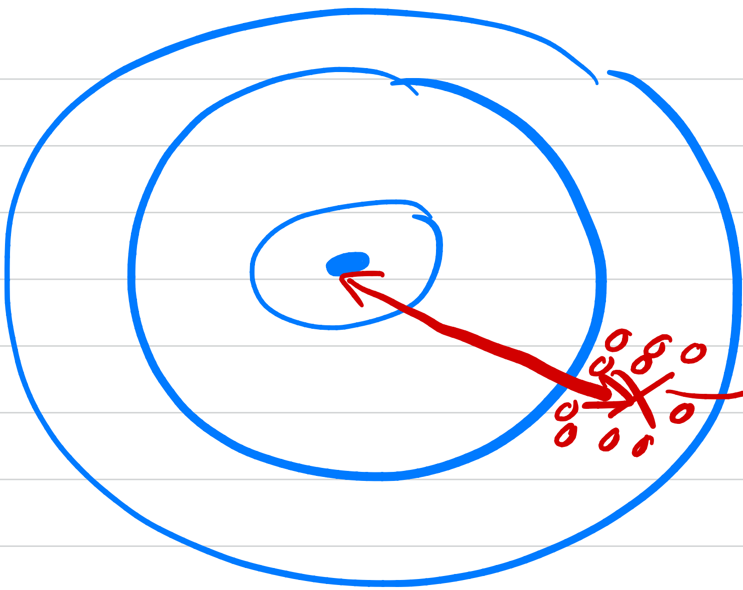
$$\underbrace{\mathbb{E}\left[\left(f-\mathbb{E}[\tilde{y}]\right)^2\right]}_{\text{Bias}} + \underbrace{\mathrm{var}[\tilde{y}]}_{\text{variance}}$$

$$+ \sigma^2$$

f is an unknown, then we replace f with the output y

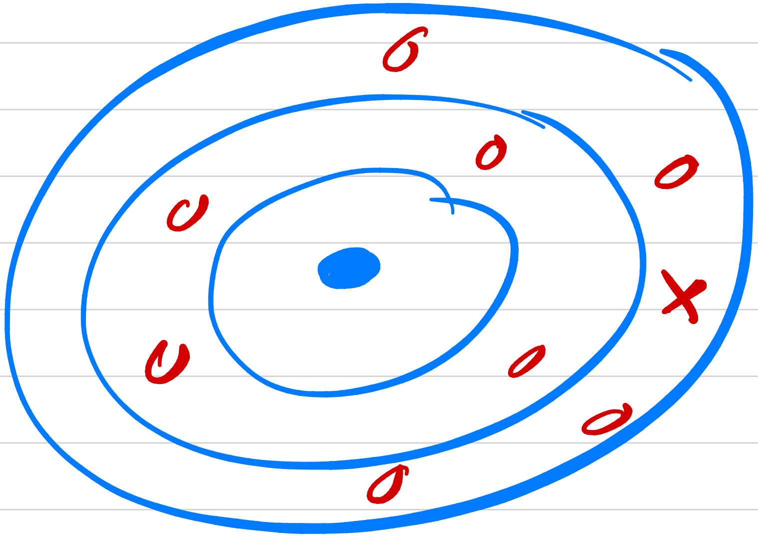
$$\text{Bias} \simeq \frac{1}{n}\sum_{i=0}^{n-1}\left(y_i - \mathbb{E}[\tilde{y}]\right)^2$$

Analysis here is made on test data

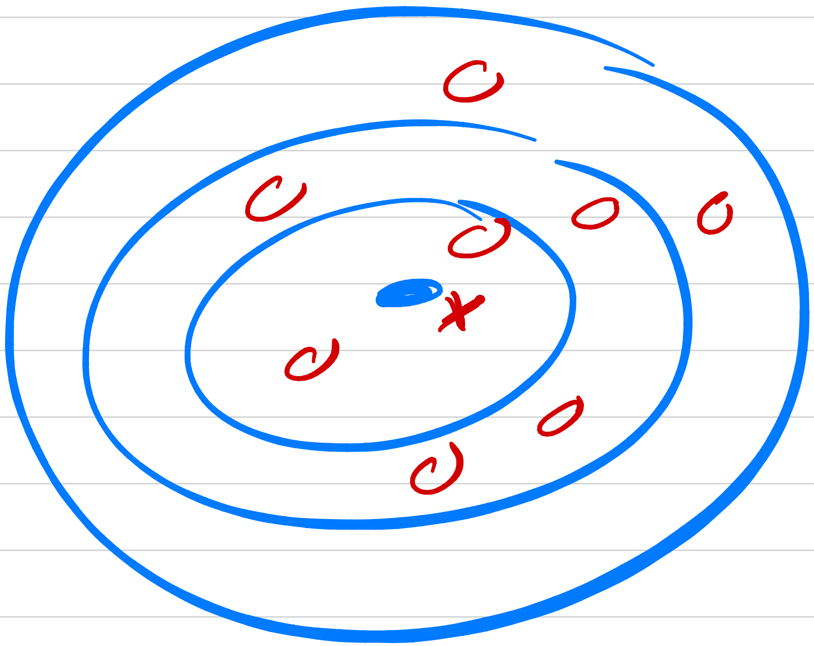


High Bias
Low variance
(low spread in model prediction)

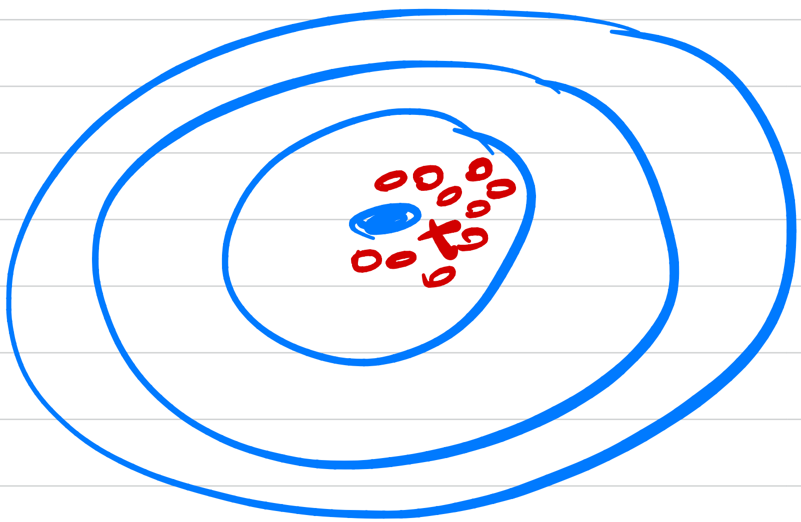
$E[\tilde{y}]$



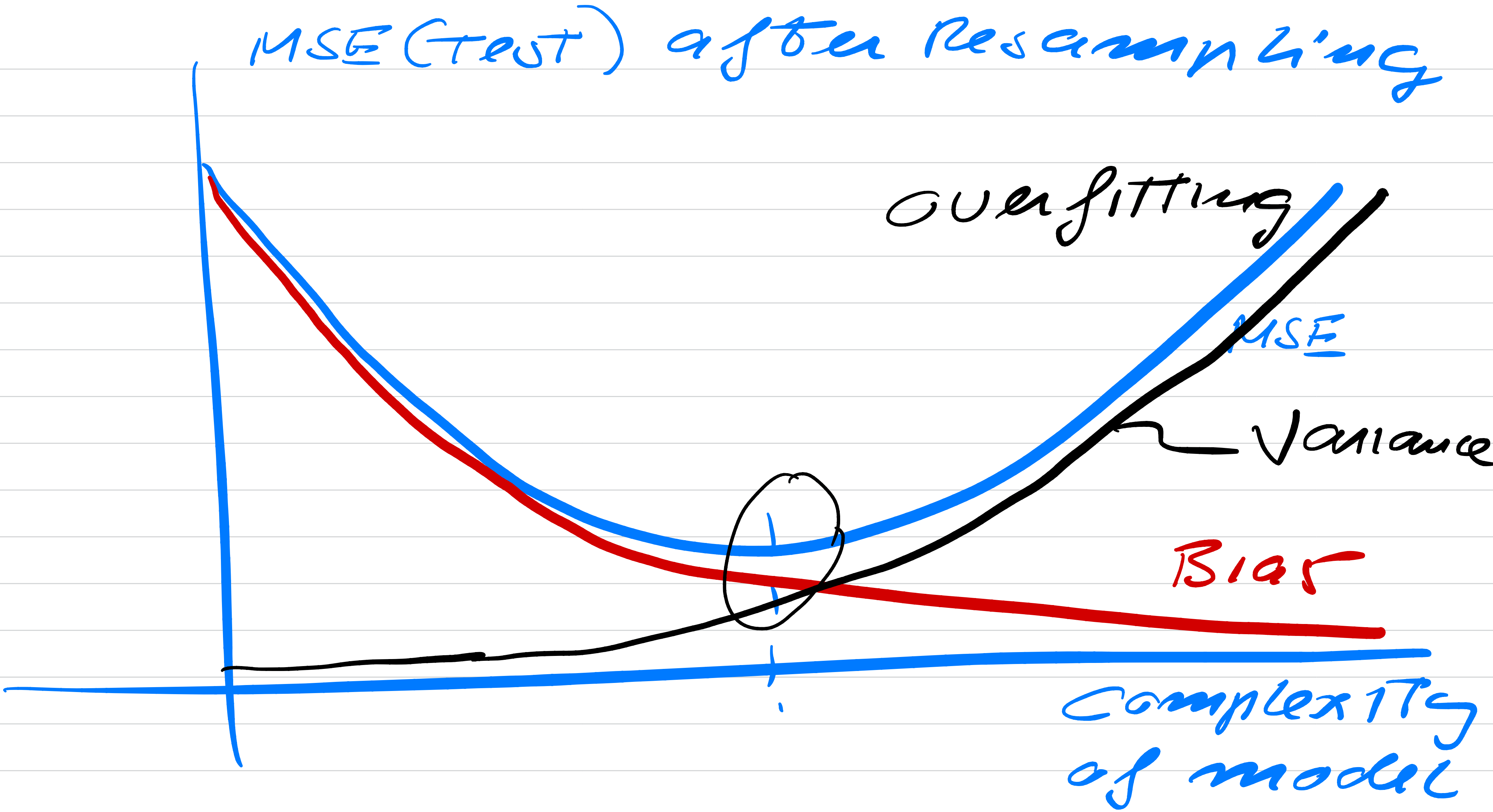
High Bias
high variance
high spread



Low Bias
high variance
(high spread)



Low Bias
Low variance
ideal case


MSE (Test) after Resampling
overfitting
MSE
Variance
Bias
Complexity
of model

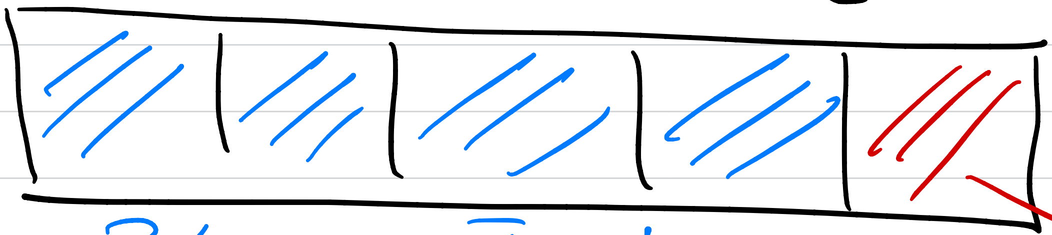
Cross-validation

Divide data in K-folds

K = 5

n data entries

(i)

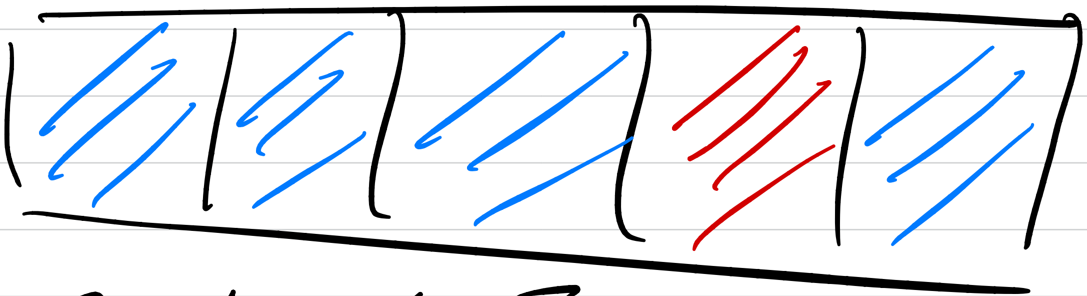


Blue = Train

Test

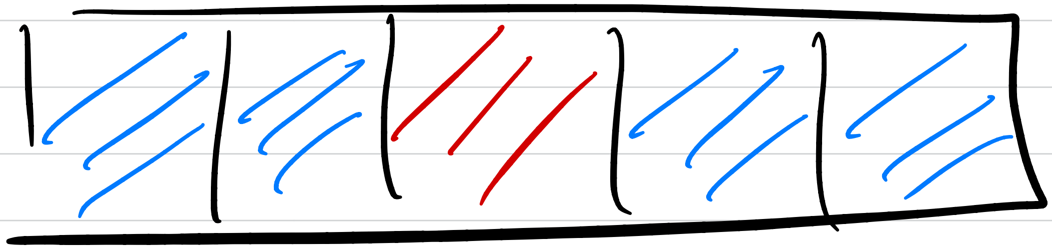
calculate MSE(Test(i))

(ii)



Calculate MSE(Test(ii))

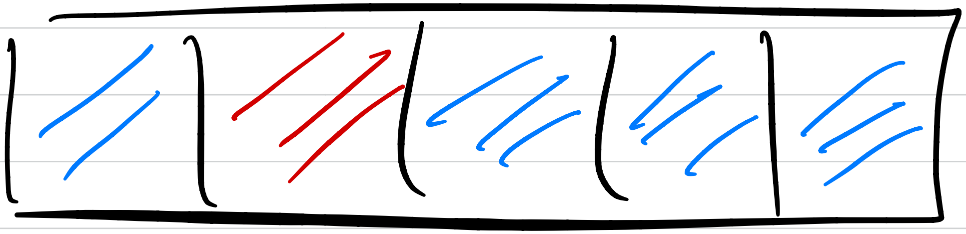
(iii)



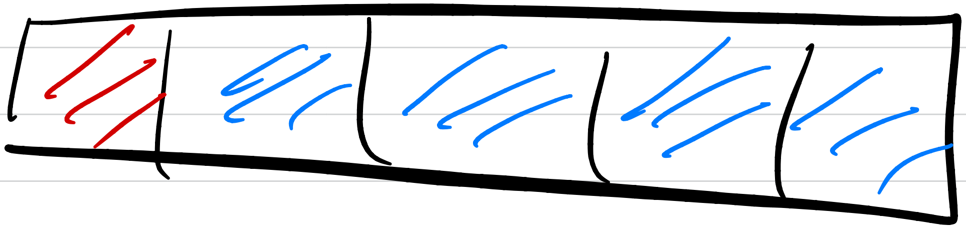
$MSE(Test(iii))$

$\boxed{K \sim 5-10}$

(iv)



(v)



$$MSE(Test) = \frac{1}{5} \sum_{i=1}^{5} MSE(Test(i))$$