Bias variance

Friday, 17 December 2021 7:54 PM

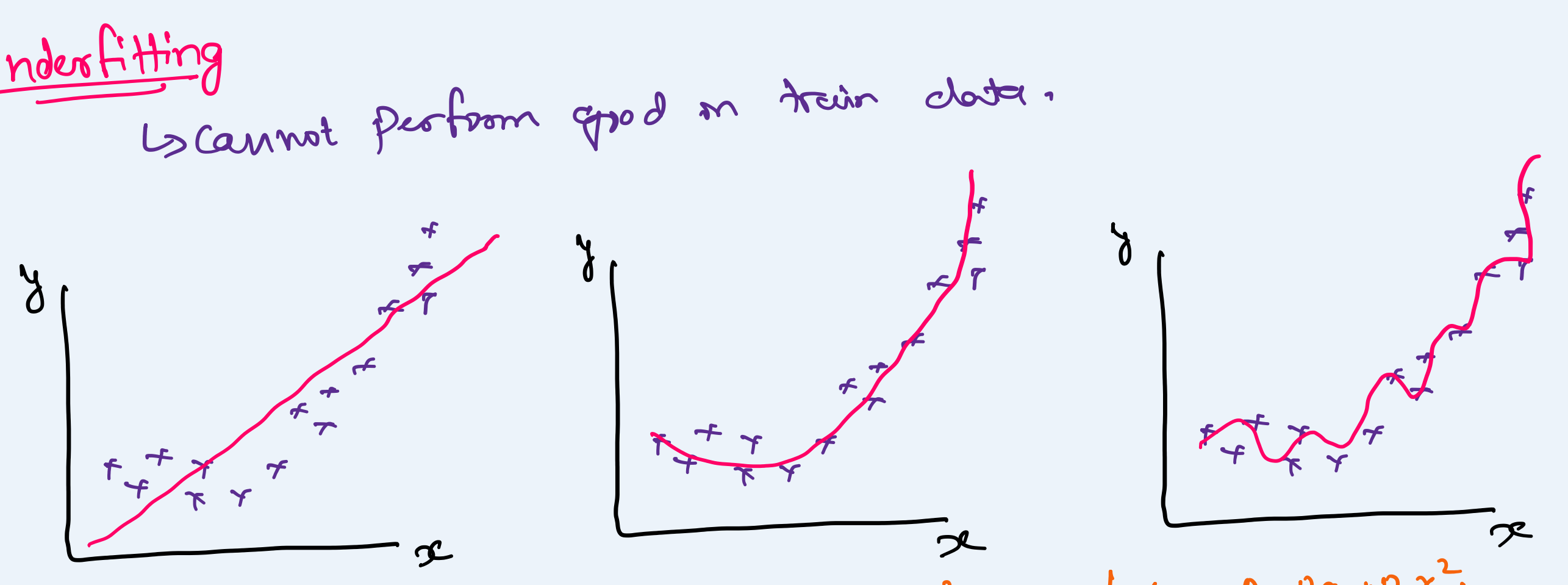
→ How to check if model is overfitting ?

Train ⇒ v.good. } → overfit 0.95

Test ⇒ v.bad 0.68

Underfitting

↳ cannot perform good on train data.



$h_\theta(x) = \theta_0 + \theta_1 x^1$

[Underfitting]

train : 70%

test : 68%

$h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2$

Generalized

[just Right]

train = 88%

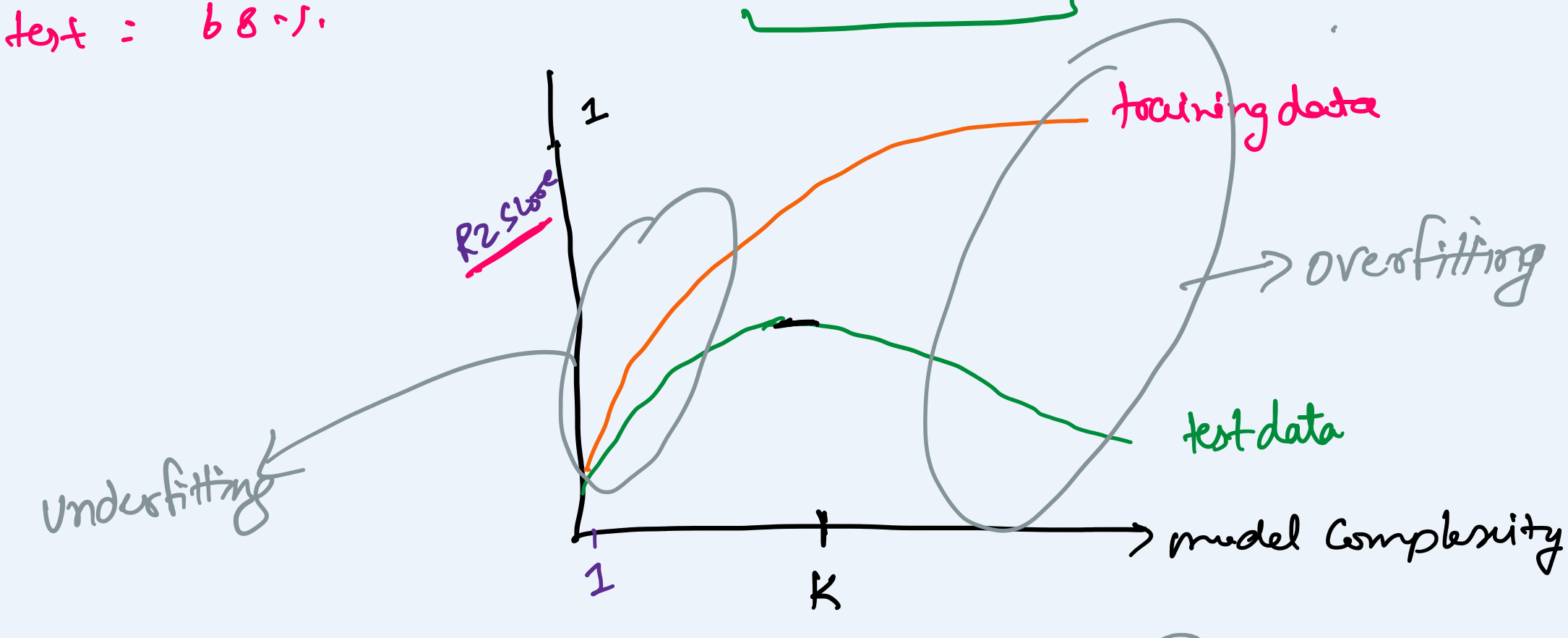
test = 85%

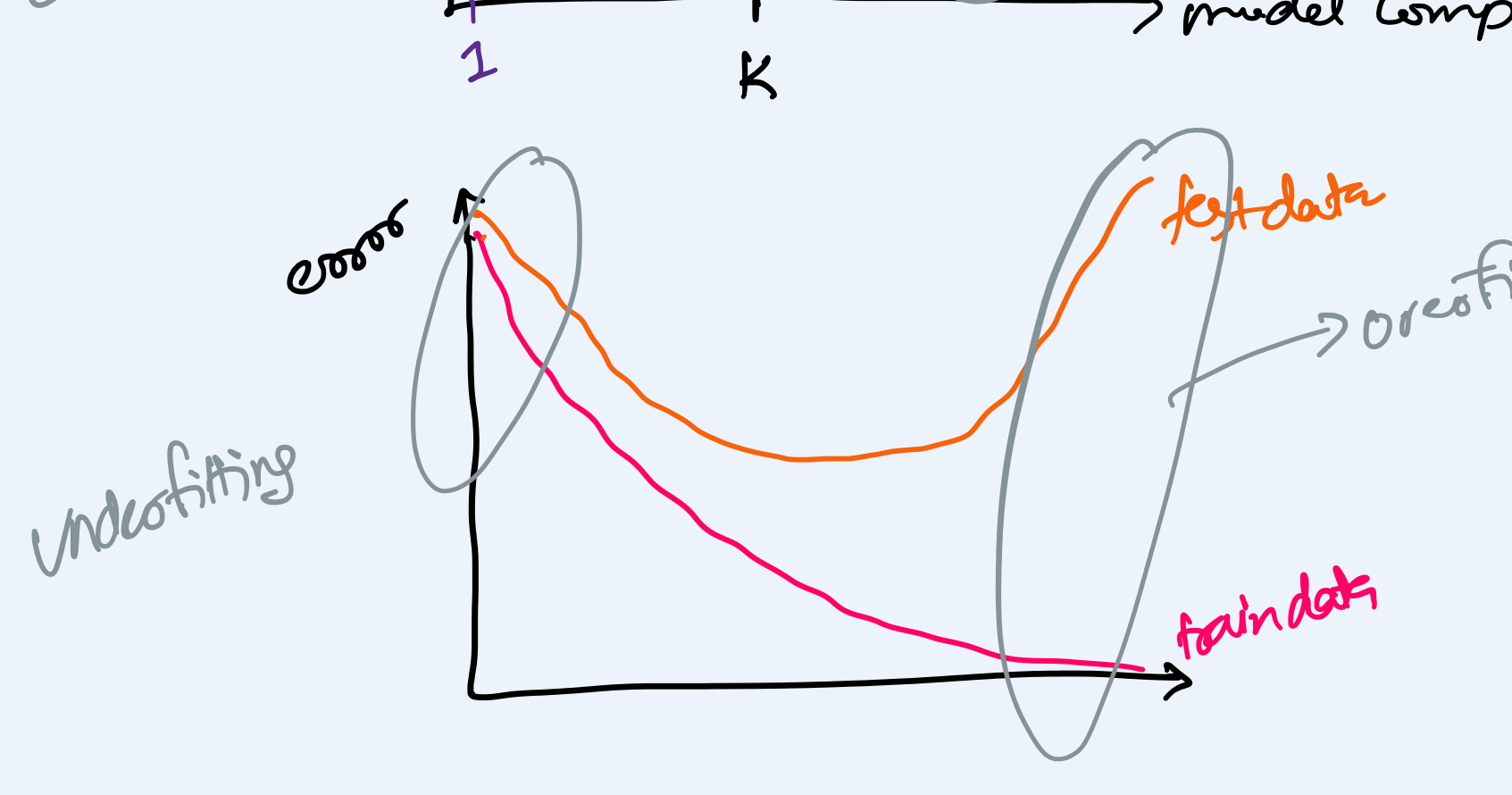
$h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \dots + \theta_{10} x^{10}$

[Overfitting]

train = 98%

test = 75%



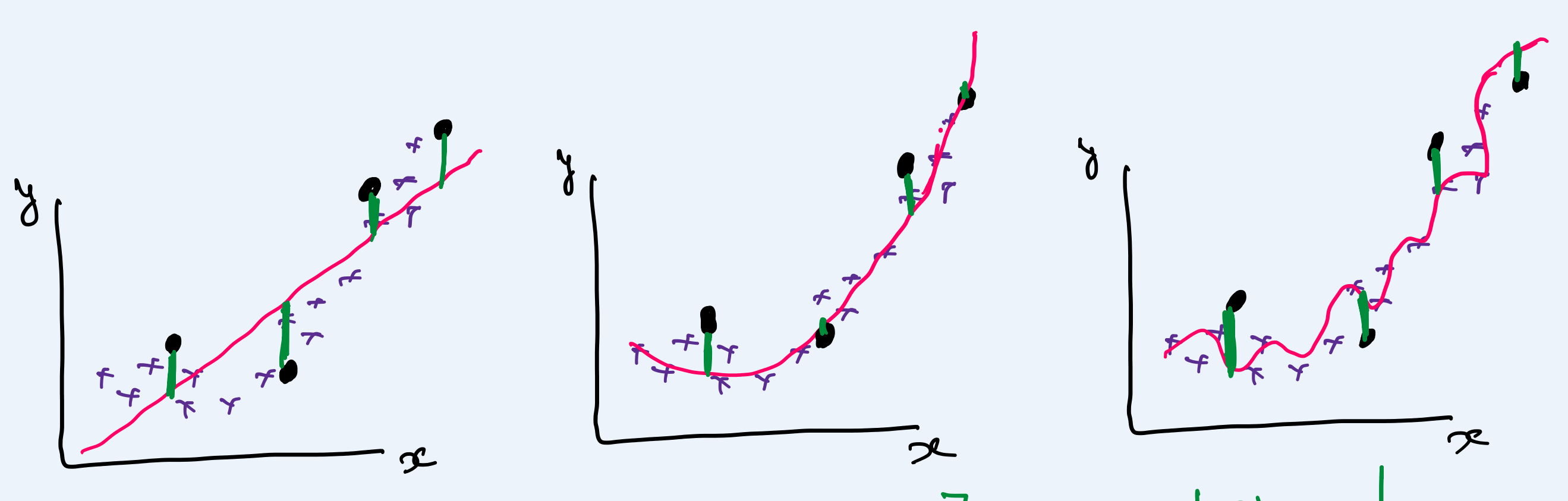


model complexity →

Bias Variance Tradeoff

① Bias → too simplified assumptions about the data

② Variance → too much assumptions about the data.



Train ↑ test ↑

Train [moderate] test low

train ↓ test ↑

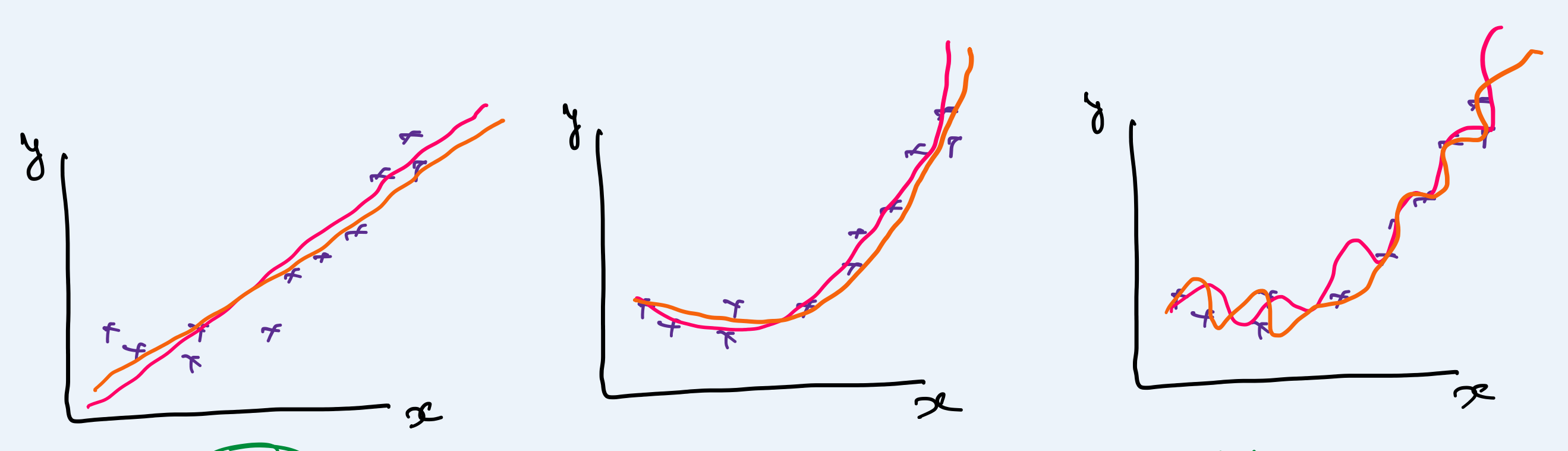
	High	Low
Bias		
Variance		

High Bias → Underfitting

High Variance → Overfitting

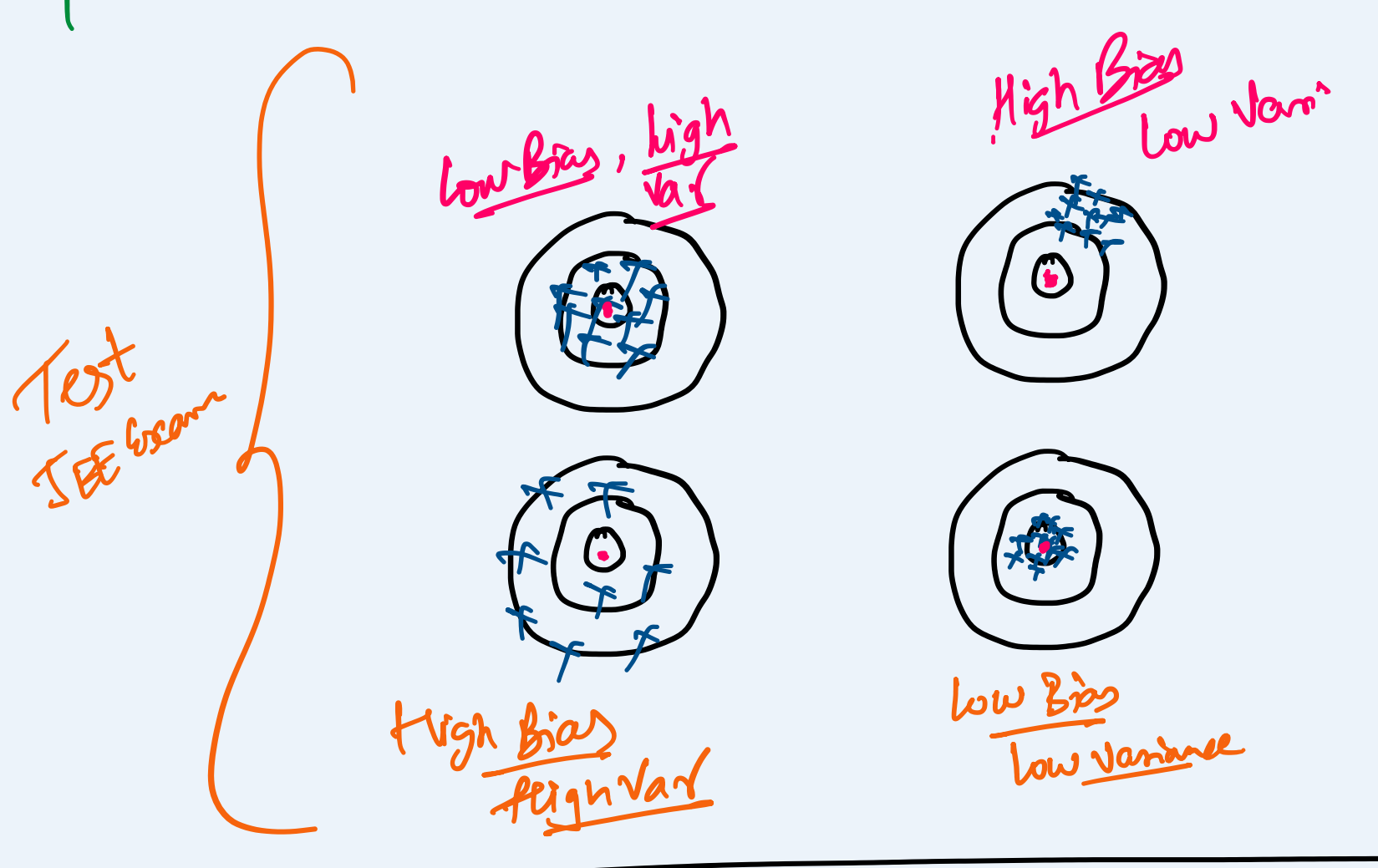
Idea Generalized

Low Bias & Low Variance



↗ High Bias

High Variance model [overfitting]



	High Var	Low Var
High Bias		
Low Bias		✓

Model

↳ Parameters : θ's (learnable Parameters) $\theta_0, \theta_1,$

↳ Hyper Parameter : Non-learnable Param. Programmer set these Param before model training.

e.g → ① learning rate : lr = 0.1, 0.01

② max_itr : 100 (80%), 200 (85%), 300 (85%)

③ degree = 1, 2, 3, 4 . 10 K ?

Model Selection

test-set → M1 → 91%, M2 → 89%, M3 → test data → 93% → Performance score 93%, M10 → 85%

Unseen score → 88%

data leakage

80% Train | 10%

60% Train Set	20% Validation set	20% Test set

Train 10 Models

degree K=1, K=2, K=3 ✓ ... K=10

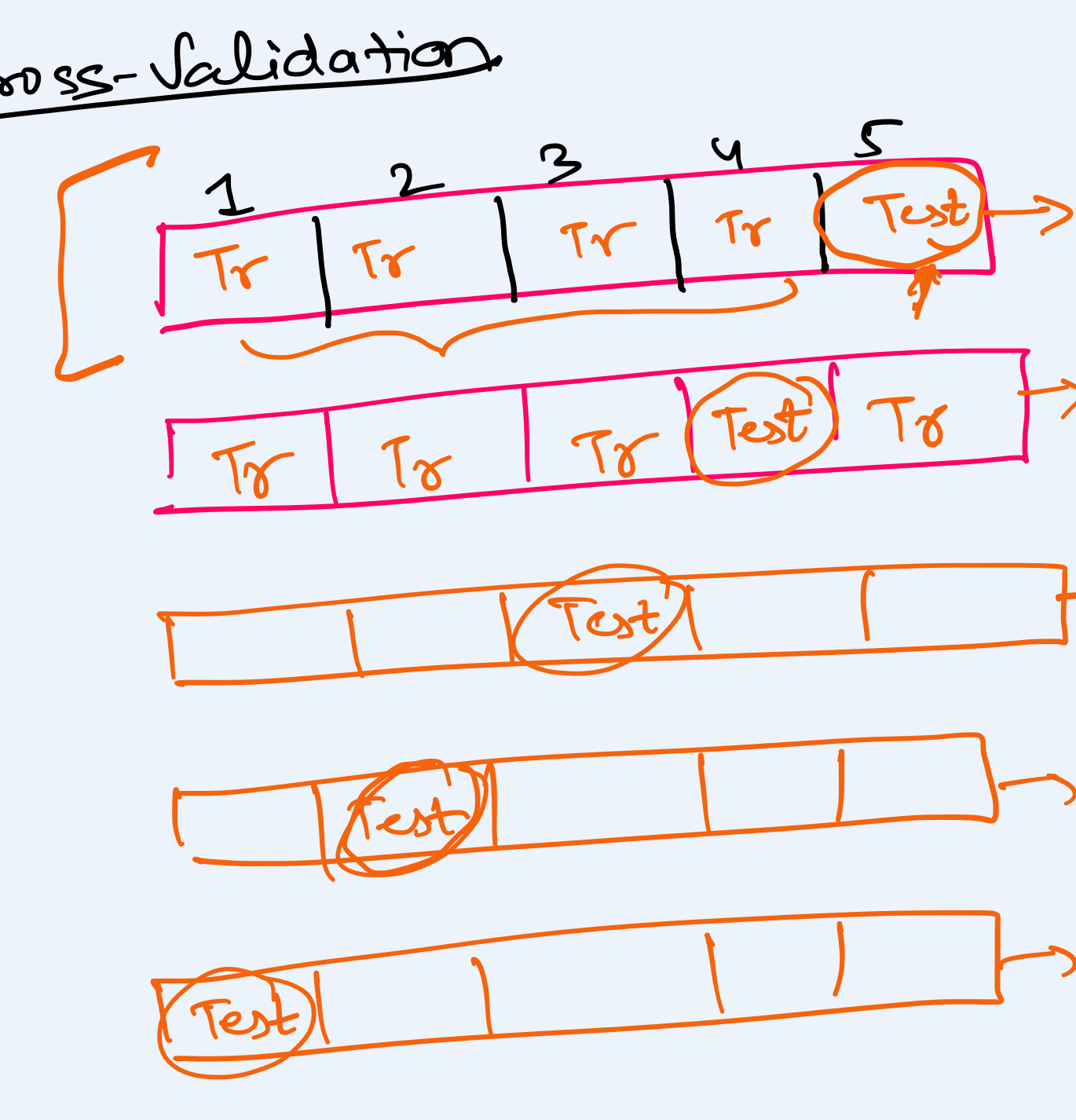
V.S → 93%

M3 → Test Data → 91%

cross-Validation

K-fold cross-Validation

K=5



→ M1 Score 1 ≈ 95%

→ M2 Score 2

→ M3 Score 3

→ M4 Score 4 → 86%

→ M5 Score 5

avg Score

Regularization (Put some restrictions)

↳ Prevents overfitting & underfitting.

↳ shrinking the Parameters.

$h_\theta(x) = \theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^3 \dots + \theta_{10} x^{10}$

↓ -0.001 ↓ 0.000001

$\theta \in (-\infty, \infty)$

$\theta_4 = \theta_5 = \dots \theta_{10} = 0$

$\theta_{10} \approx 50$ (x_{10})

$\theta_2 = 5$ (x_2) ← 10 features →

Error | loss = $Loss_{model} + L_{regularization}$

$L_{regular} = \sum_{i=1}^{n} |\theta_i|^2$ L2 regularization

$L_{regularization} = \sum_{i=1}^{n} |\theta_i|$ L1 regularization

L2 → Ridge Regression

$Loss = \frac{1}{2m} \sum_{i=1}^{m} (y^{(i)} - \hat{y}^{(i)})^2 + \lambda_1 \sum_{i=1}^{n} \theta_i^2$ (L2 regular.)

L1

$Loss = \frac{1}{2m} \sum_{i=1}^{m} (y^{(i)} - \hat{y}^{(i)})^2 + \lambda_2 \sum_{i=1}^{n} |\theta_i|$ (L1 regular.)

↳ Lasso Regression

ElasticNet Reg = MSELoss + L1 + L2

λ ↑ e.g λ = 10000 Underfit

$\theta_1 = \theta_2 = \theta_3 = = \dots \theta_n \approx 0.001$

$h_\theta(x) = \theta_0$ Param are shrinking.

λ ↓ e.g λ = 1

↳ overfit.

$h_\theta(x) = \theta_0 + \theta_1 x + \theta_1 x^2 + \theta_3 x^3 +$

(Complex)