GO CLASSES

Overfitting and Underfitting

Mohd Ayyoob
(M.Tech CSA, IISc Banglore)

2024

AGENDA

1. Overfitting
2. Underfitting

} Why do we even need it?

3. Bias
4. Variance

} 'Hand-wavy' introduction

5. Code-snippets and Visualizations.

GOAL in Machine learning:

1. We want a model to 'generalize' well to unseen data (TEST.) data.
2. We want our model to have high/low generaliz^{n} accuracy.
3. We want our model to have high/(low) generaliz^{n} error

$$\theta^{*} = \arg\min_{\theta} \underbrace{\frac{1}{2}(\hat{y} - y)^2}_{\text{MSE}} = \arg\min_{\theta} \frac{1}{2}(h_\theta(x) - y)^2 + \|w\|^2$$

TRAIN-ERROR = 0

TEST-ERROR ↓↓↓

M1

E_{train} ↑

M2

E_{test} ↑↑

M3

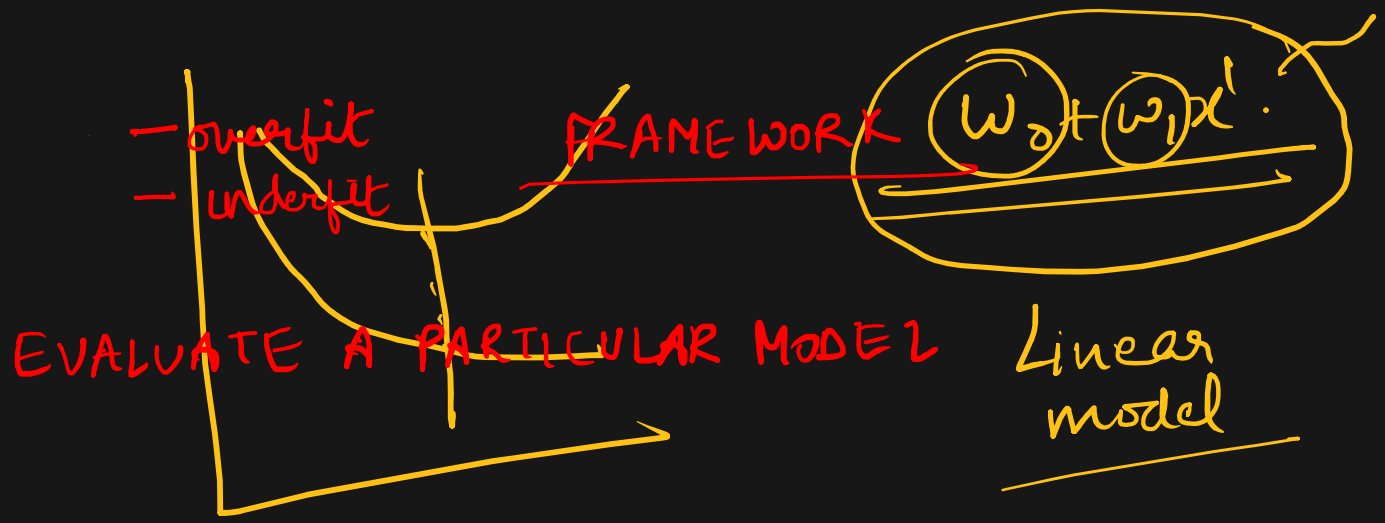
E_{test} ↓

E_{Train} ↑

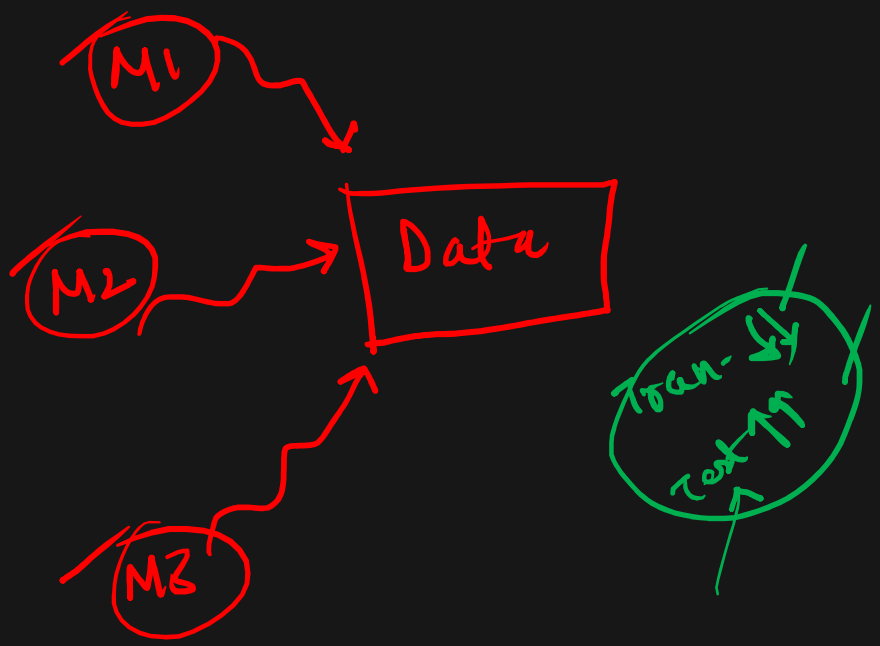
M4

A_{train} ↑

A_{test} ↓



Linear model



— TEST-error ↓↓ ✓

FRAMEWORK

** compare different models on same D.

① No of param

M1 : degree 1 : LR —
M2 : degree 5 : Poly LR —

M2 > M1

M2 is complex > M1

② underlying algo :

M1 : LR
M2 : SVM
M3 : NN

No of parameter ↓

M1 > M2 > M3

$$[\text{TEST-ERROR} = \underline{\text{BIAS}^2} + \underline{\text{VARIANCE}} + \underbrace{\sigma^2}_{\text{irreducible error}}]$$

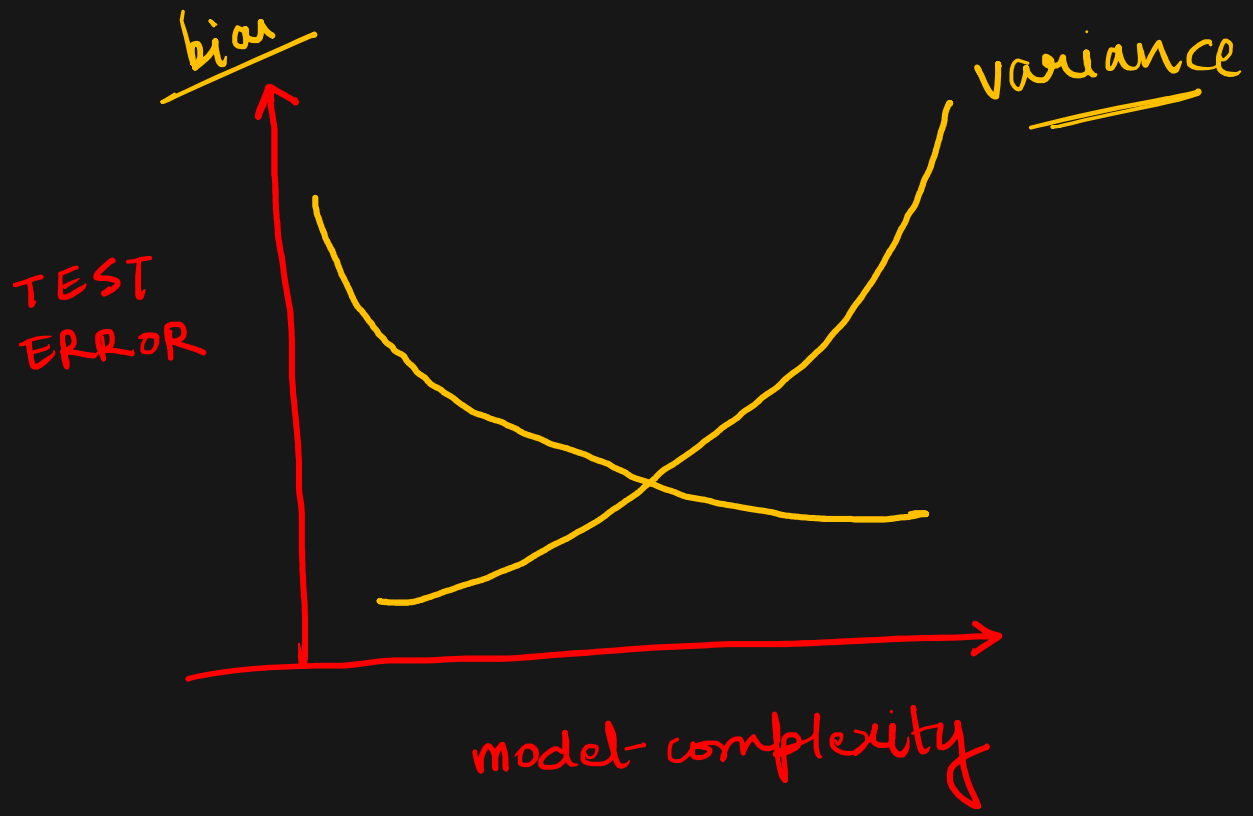
How good my assumed structure actually representing the truth

model is too sensitive to data

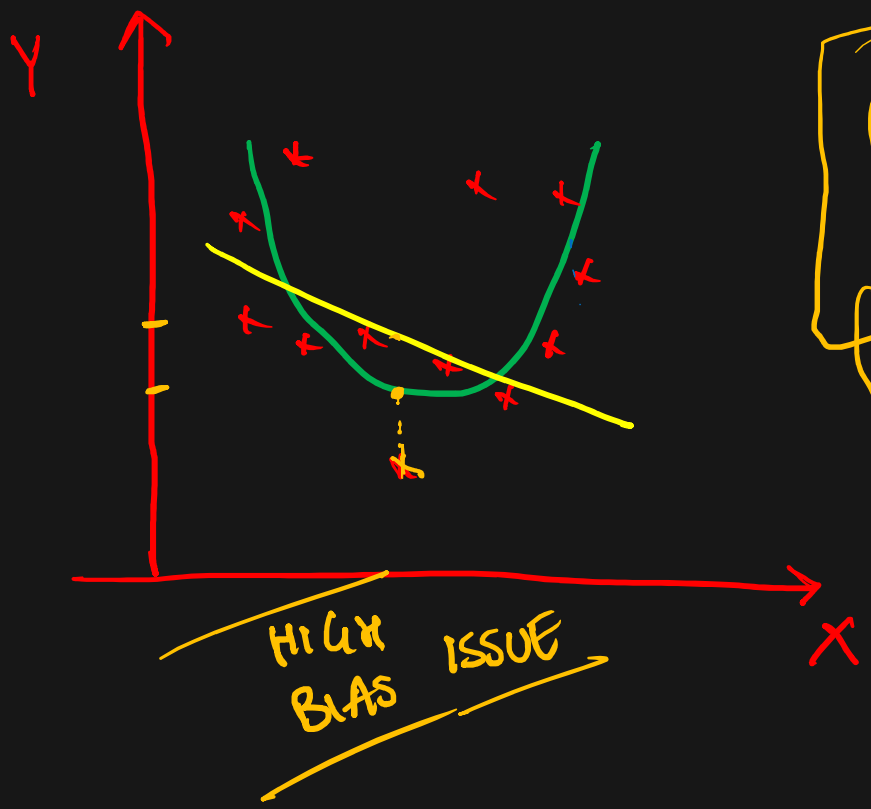
M1 : High Bias

M2 : High Variance

BIAS- VARIANCE TRADEOff



UNDER-FITTING



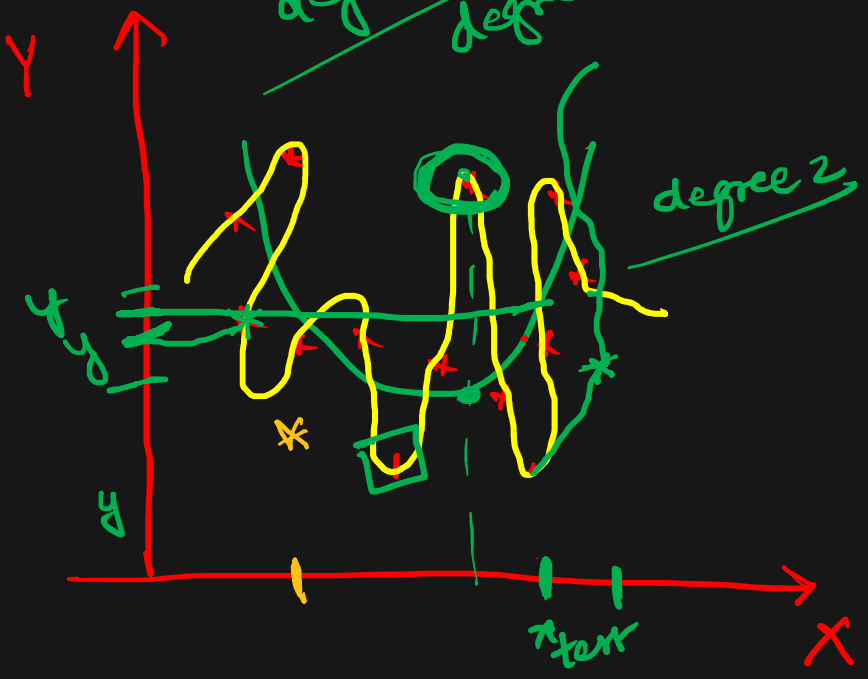
– : true func^{n} → train-error : ↑↑
– : M1 test-error : ↑↑

Assumed : model that we are fitting is from linear family WTF

→ model is too simple ←
→ make model more complex

BIAS

OVERFITTING



- : true func^n.
- : M_2

- train-error : 0 ↓↓
- test-error : ↑↑

model is too complex —

- model is fitting noises ↑
- model is too sensitive to changing training data. ↗

HIGH VARIANCE

→ Adding more points would help in this

variance of model's prediction

GOOD - FIT

— : true func^{n}.
— : M3

train-error :
test-error :