Overfitting And Underfitting (Bias And Variance)

1. Training dataset
2. Test dataset
3. Validation dataset

70%-30%

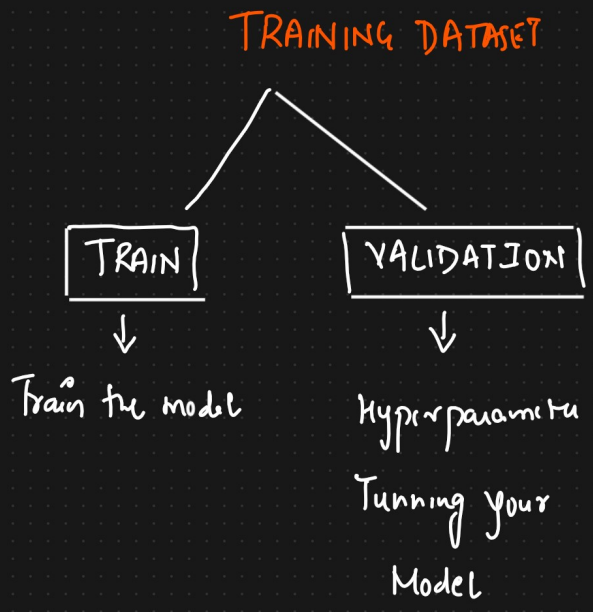
Dataset

1000 datapoints

→ TRAINING DATASET (700) → Model To TRAIN

→ TEST DATASET (300) → Testing Model

Size of House	No. of bedrooms	Price
—	—	—
—	—	—
—	—	—
—	—	—



Underfitting :-

- Poor performance on both training and test data.
- High bias (makes strong assumptions about the data)

Overfitting :-

- High performance on training data but poor performance on test data.
- High variance (model is too sensitive to small changes in data).

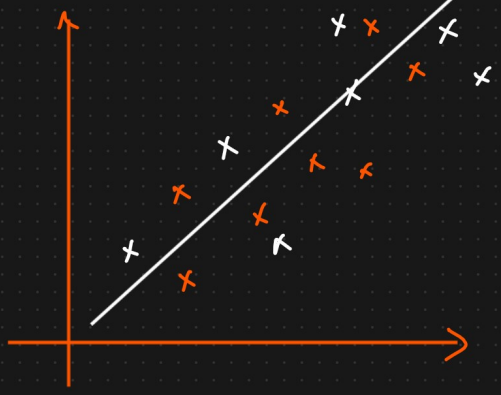
TRAIN	Very Good Accuracy (90%) [Low Bias]	Very Good Accuracy (90%) [Low Bias]
TEST	Very Good Accuracy (85%) [Low Variance]	Bad Accuracy (50%) [High Variance]
	⇑ → Generalized Model	⇓ Model is Overfitting

TRAIN — Model Accuracy is low [High Bias]

TEST — Model Accuracy is low [High Variance]

⇓

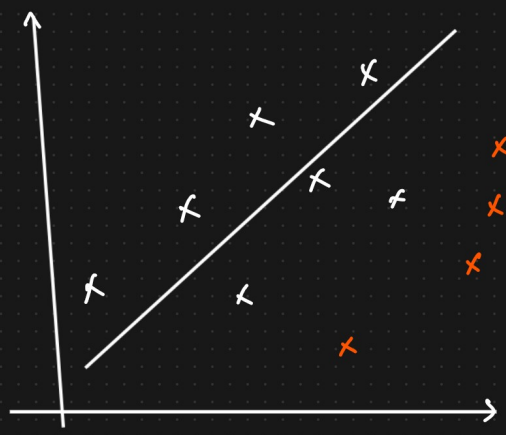
Model is Underfitting



Generalized Model

⇓

Low Bias, Low Variance



Overfitting

Low Bias, High Variance

- Bias :- Bias is the error due to overly simplistic assumptions in the model
- Variance :- Variance is the error due to excessive sensitivity to the training data