

- ① Train-Test split
- ② Model training and performance
- ③ Overfitting and Underfitting
- ④ Bias and Variance

train → model training

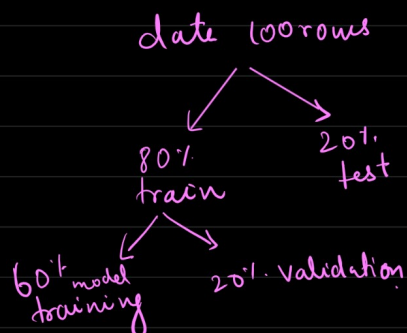
# Area of house	# of rooms	y (Price of house)
1100	2	2.1
—	—	—
—	—	—
—	—	—

learn → Pattern → $y = f(x)$

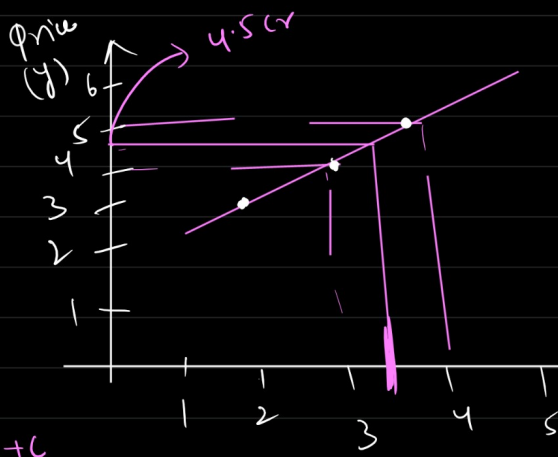
↑ input feature

↑ Output (Price)

↑ predictive function



* Model training : Given the data estimate the prediction function by minimizing the error.



# of rooms	Price of house in cr
2	3
3	4
4	5
3.5	? (4.5 cr)

$$y = mx + c$$

Date # of room

Price of house in cr

data

```

graph TD
    A[70% : 30] --> B[80% : 20]
    A --> C[90% : 10]
  
```

train 70%

testing 30%

3
4
5
6
3.5
8.5
9.5
1.2

In this 70% you will train the model

↓

Predictive function

* Model performance good on test data \uparrow
(Accuracy should be high)

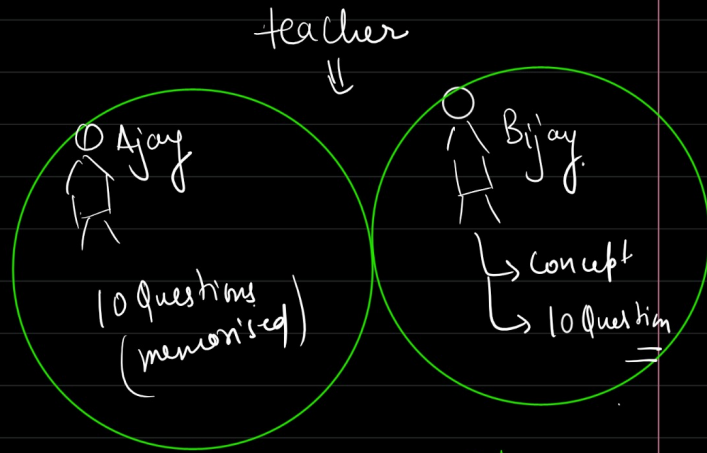
* Model Overfitting

train data \rightarrow Model is trained \rightarrow Accuracy \uparrow
is high (95%)

test \rightarrow Model is tested \rightarrow Accuracy is low
(unseen data) (65%)

\rightarrow Overfitting

* Model performs well on train data but worst on test data.



testing \rightarrow Board / Exam

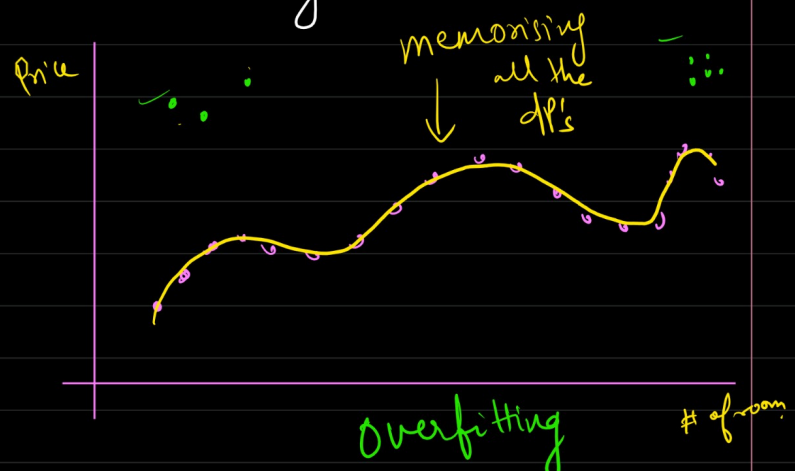
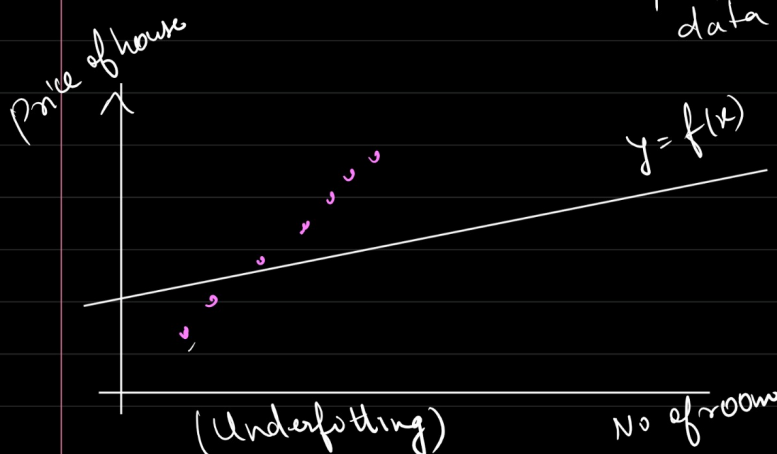
Questions are slightly changed.

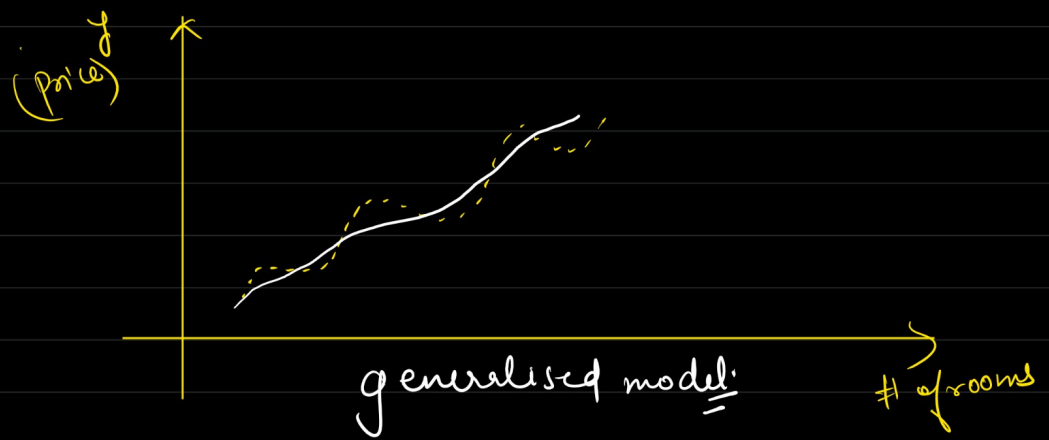
* Model Underfitting

train data \rightarrow Accuracy $\downarrow \rightarrow 50\%$
test data \rightarrow Accuracy $\downarrow \rightarrow 40\%$ } Underfitting

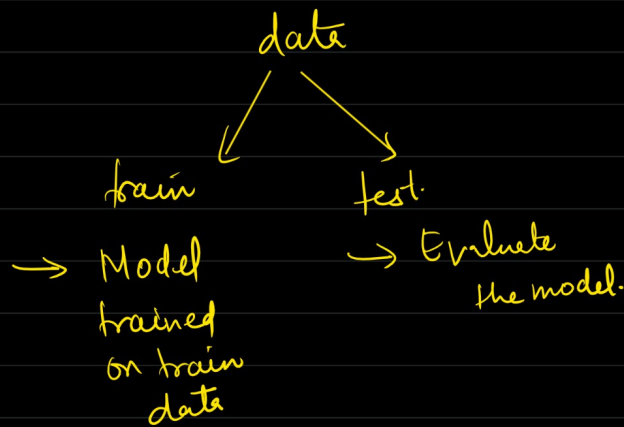
* Generalised model

train data \rightarrow Accuracy \uparrow 85%
test data \rightarrow Accuracy \uparrow 84%





Bias / Variance



Model training

# of rooms	Price of house (y) in Cr.	$y = f(x)$ train	Error
2	3	2.9	3 - 2.9
3	4	4.2	4 - 4.2
4	5	6.1	-
5	6	-	-
7	8.5	-	-
8	9.5	-	-
9	11.5	-	-
3	1.2	5	1.2 - 5
9	3.5	6	3.5 - 6
10	6.2	2	6.2 - 2

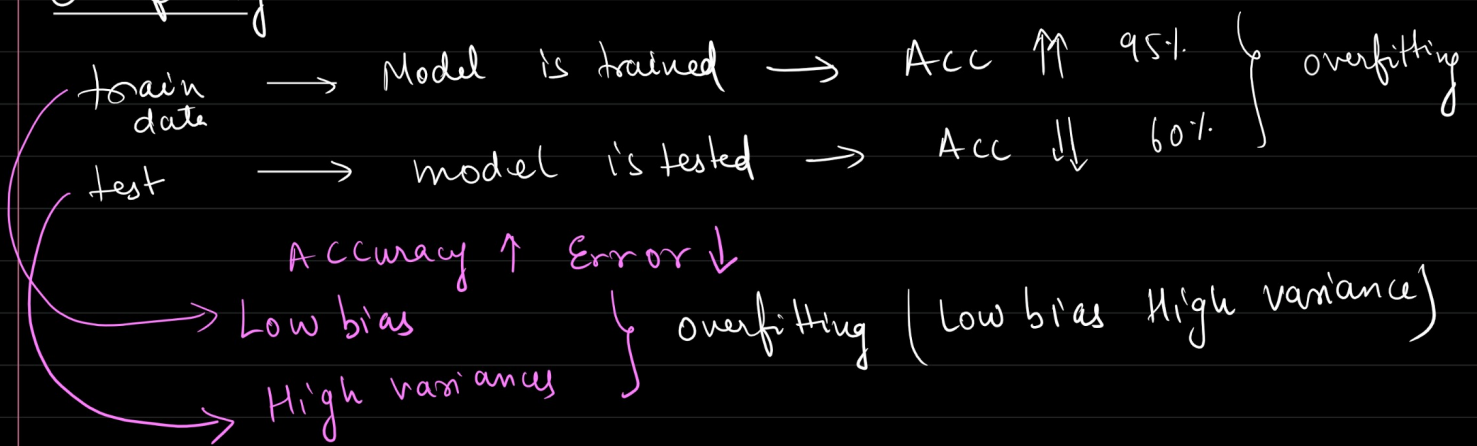
→ training error

→ testing error

* training error is also known as bias.
High training error mean high bias.

* testing error is also known as Variance.
High testing error means high variance.

Overfitting

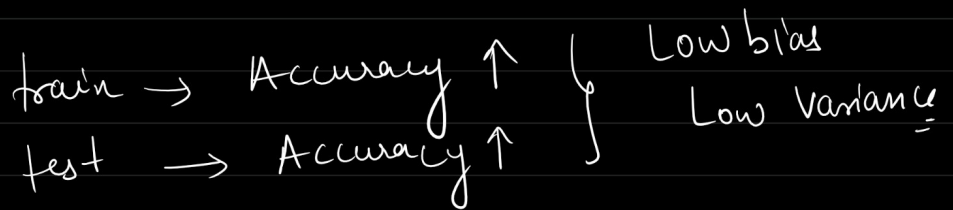


Underfitting



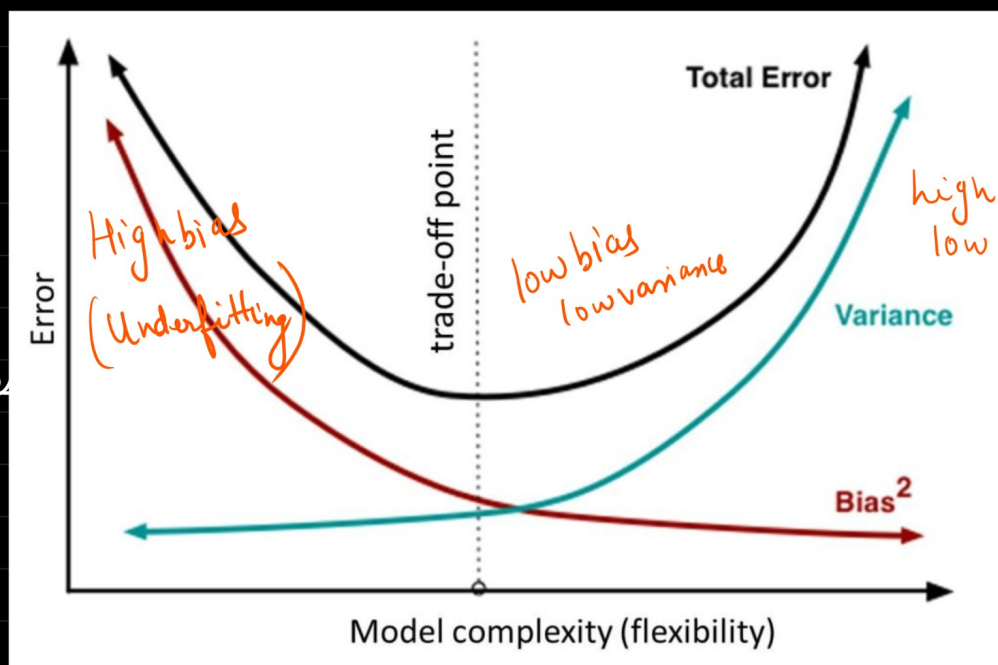
Underfitting $\left\{ \begin{array}{l} \text{High bias} \\ \text{High variance} \end{array} \right.$

Generalised Model



Bias-Variance trade off

Underfitting
 \rightarrow Some more training
dp
 \rightarrow More feature
 \rightarrow better Algorithm
to learn patterns



high variance
low bias
(overfitting)
 \downarrow
 \rightarrow Select relevant feature
 \rightarrow Don't learn from noise
 \rightarrow Use some other Algo