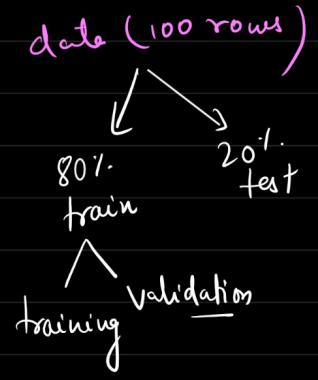
① train - test split
② Model training and performance
③ Overfitting and underfitting
④ Bias and Variance.

train — model training.



#Area of House (f1)	# of room f2	y (price of House)
1100	2	2.1
—	—	—
—	—	—
—	—	—
1200	3	?

$y = f(X)$ → input feature.

y: Output; f: Predictive fn

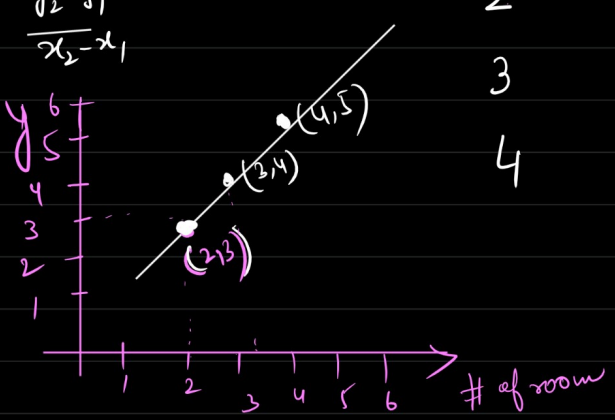
* Model training → given the data estimate the prediction function by minimising the error.

eq' of line ⇒ $y = mx + c$

$y = mx + c$

$\frac{y_2 - y_1}{x_2 - x_1}$

#of rooms	Price of house (y) (in cr)
2	3
3	4
4	5



train - test

↓ training

↓ Predictive fn

70 : 30 (train test)

80 : 20

90 : 10

# of rooms	Price of house (y) (in cr)
2	3
3	4
4	5
5	6
6	3.5
7	8.5
8	9.5
3	1.2
9	3.5
10	4.5

train: 70% (2 to 8) → In this 70% you will train the model → Predictive fn.

test: 30% (3, 9, 10)

* Model performance ↑ (Accuracy should be high) Viswa sir (final)

* Overfitting

train → Model is trained → Accuracy is high ↑↑ (95%)

test (unseen) → Model is test → Accuracy is low (65%)

10 Questions memorises

Concept

10 Question

→ Overfitting

* Model performs well on train data, but wose on test data.

* Underfitting

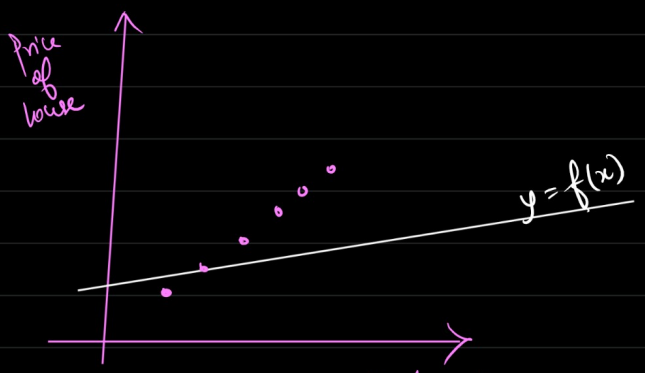
Train → Accuracy ↓ → 50% } Underfitting

Test → Accuracy ↓ → 40% }

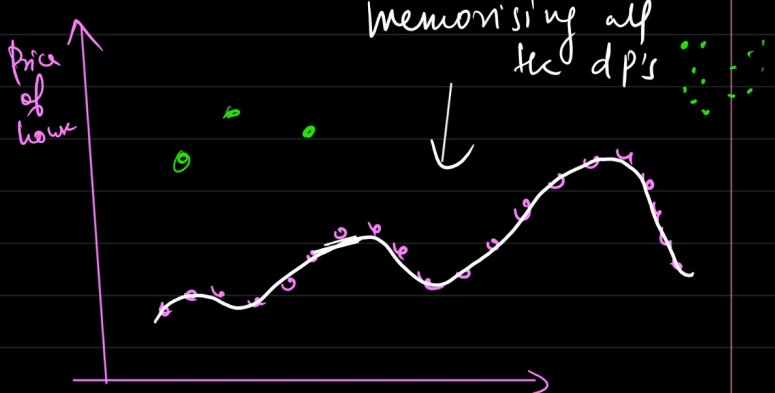
* Generalised model

train → Accuracy ↑↑ = 85%

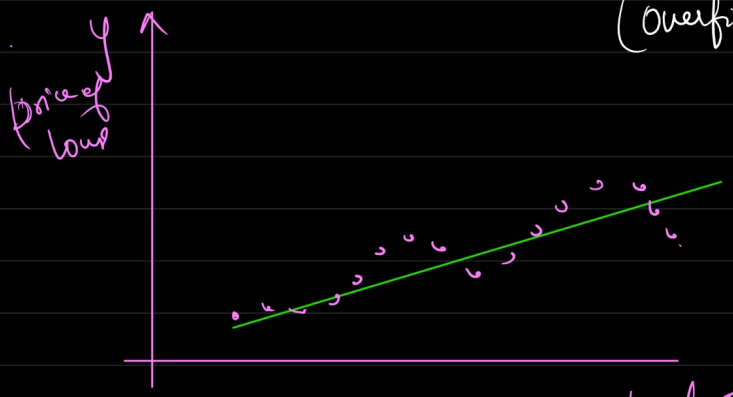
test → Accuracy ↑↑ = 84%



(Underfitting)



(overfitting)



(generalised model)

* Bias | variance

→ Model trained on trained

# of rooms	Price of house y (in cr)	$y = f(x)$ train	y_{train} Prediction	Errors
2	3		2.9	3 - 2.9
3	4		4.1	4 - 4.1
4	5		3.2	5 - 3.2
5	6		6.1	
6	3.5		2.3	
7	8.5		2.2	
8	9.5		3.5	
3	1.2		3	1.2 - 3
9	3.5		4	3.5 - 4
10	...		5	4 - 5 - 5

training error

testing error

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

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

Overfitting

Train → Model is trained → Accuracy ↑↑ 95%
Test → Model is tested → Accuracy ↓↓ 60%
} Overfitting

Accuracy ↑ error ↓

→ Low bias
→ High variance
} Overfitting
{Low bias, High variance}

Underfitting

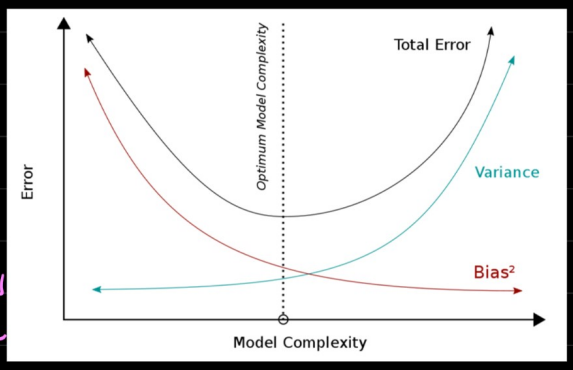
train → Accuracy ↓↓ → High training error → High bias
test → Accuracy ↓↓ → High testing error → High variance.

Underfitting } High bias, high variance

* Generalised model

train → Accuracy ↑
test → Accuracy ↑
} Low bias, Low variance.

Bias / variance trade off.



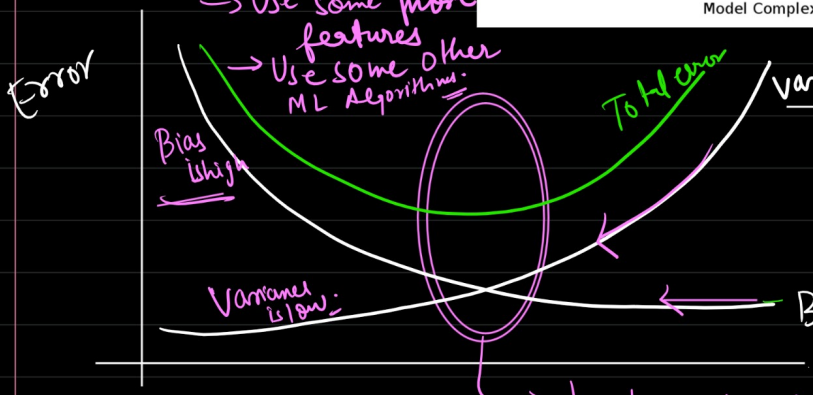
Underfitting

→ Bias is high } → Var is high

→ Some more training dp's

→ Use some more features

→ Use some other ML Algorithms.



{ high variance, low bias }

Overfitting

→ Select relevant feature.

→ Don't learn noise

→ Use some Algorithms.