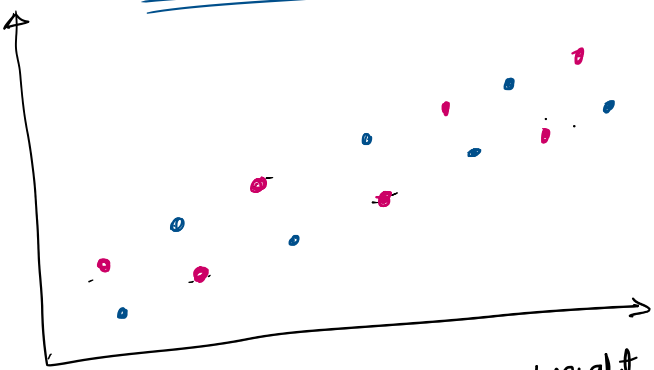
Total given Datapoints

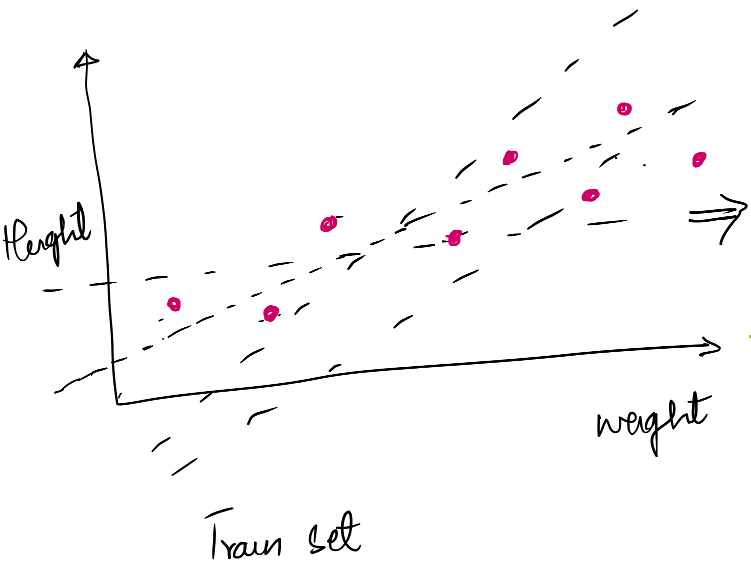
Consider →

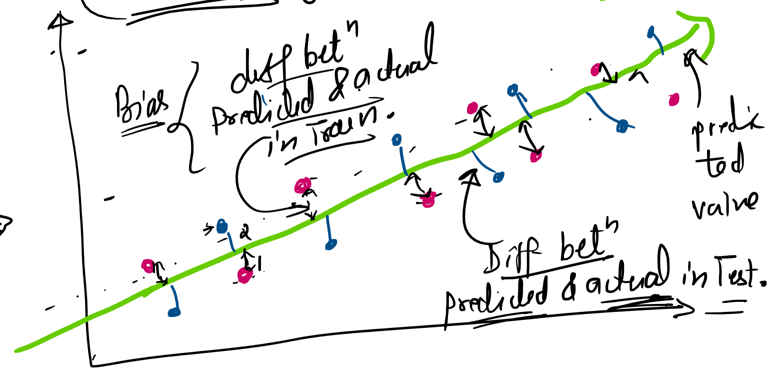


bias ↑ variance ↓

(Underfitting)

model be consider straight line



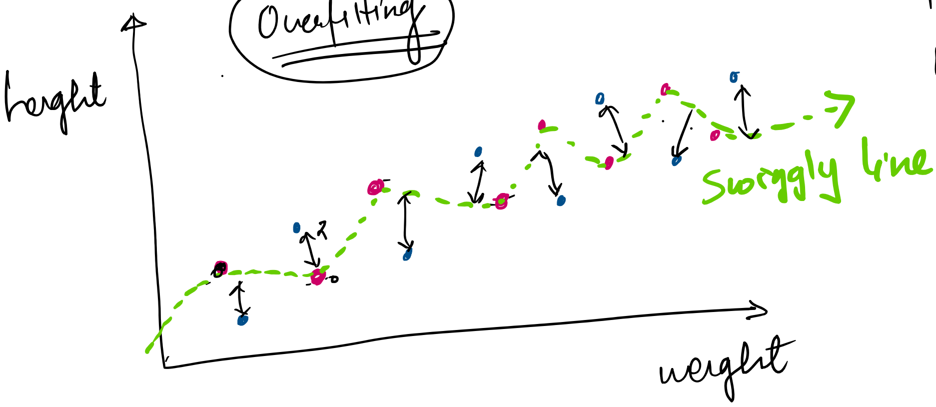


$\boxed{\text{Bias}}$ → Inability of ML methods (linear reg) to capture true relationship

Here any simple straight line cannot match all data points in Training set

bias ↓ variance ↑

(Overfitting)



Here the swiggly line handles the true relationship bet^n weight & height

So since the line almost passes through every training data points we say ⇒ Little/Zero

. Data bet^n

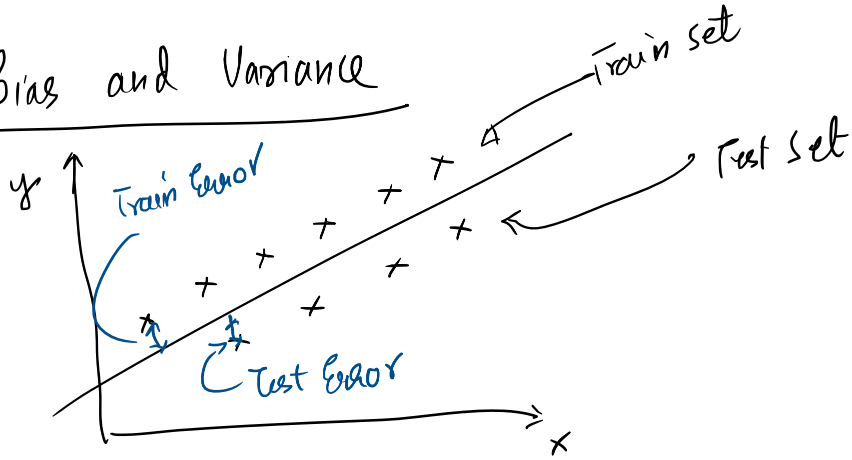
points we say ⇒ [illegible]

Zero

Bias.

Variance ⇒ Difference in fits betn the Datasets (Train & Test Data Sets).

Bias and Variance



Straight Line
↳ Polynomial degree = 1.

* Underfitting ⇒ When the model is Simple (lower degree Polynomial), the model might not fit the train set data points.

So Here the difference betn Actual and predicted value for Train set is Higher

This is bias.

So In Underfitting bias ↑

There will Error for predicted value and actual value for Test set

So the diff in Error for test and Train set is low

This is Variance

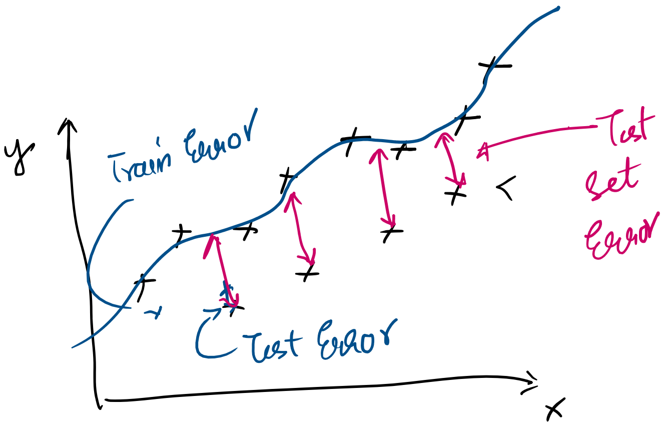
So In Underfitting variance ↓

* Underfitting

∴ For Simple Model (lower order Polynomial)

Bias ↑ Variance ↓

Bias ↑ Variance ↓



Here model is complex (higher order polynomial)

→ So Error with Train Set is very less

∴ Bias is ↓

Since the model perfectly fits the train set

It is case of Overfitting

and there is significant diff (Error) with Test set

∴ Variance is ↑

* Overfitting →

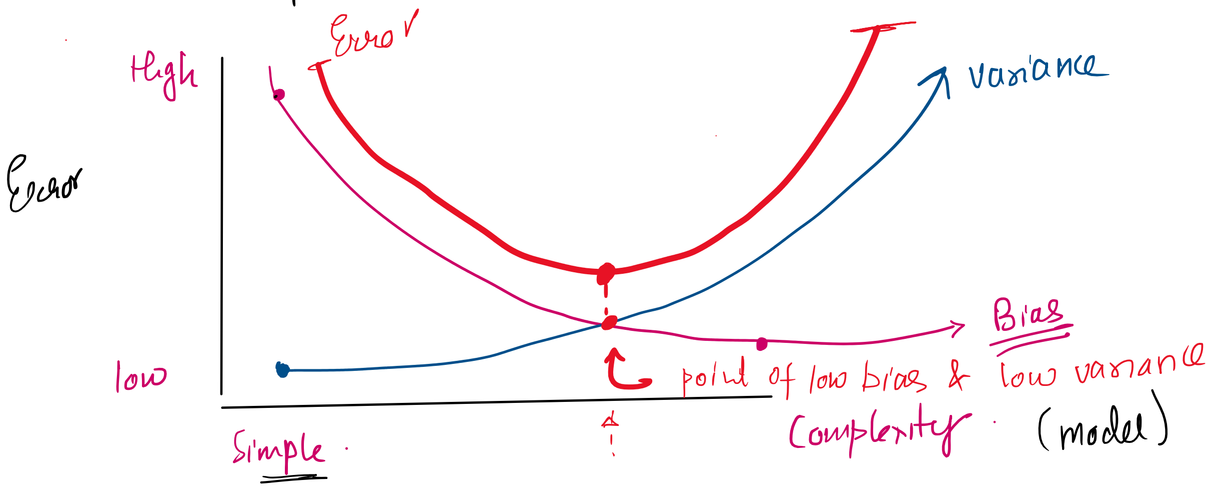
For Complex model (higher order polynomial)

bais ↓ variance ↑

Imp Bias - Variance Tradeoff

(1) If model is too simple and very few parameter → Underfitting
→ High bias
→ low variance

(2) If model is complex and has large no of parameters → Overfitting
→ Low Bias
→ High Variance.



Bias Variance Tradeoff says

We need a model that gives
(1) low bias
(2) low variance.

ie we need to find point of low bias and low variance.

The methods to achieve this :⇒

* Regularization (Penalize 'θ' parameter)
* Boosting
* Bagging.

*. to Minimize the Total Error : $Bias^2$ + Variance + Irreducible Error.

We need to Minimize the Total Error: Bias + Variance + Irreducible Error.

$$\hat{y} = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

* Performance Metric :

Consider a Skewed Data set (One Sided).

Ex

1000 MRI Scan → 990 No cancer

1000 MRI Scan → 10 cancer

there only 1% scan shows cancer

it is very less.

Normally Allowed.

→ might be Ignored

→ But this is incorrect.

In Such Case to Evaluate Performance We may use

Confusion Matrix →

		Actual: True	Actual: False
Predicted	True	True +ve (TP)	False +ve
	False	False -ve	True -ve

The above confusion matrix is OK for 2 classes.

Question → Cancer ?

		Actual: Has Cancer	Actual: Do not have Cancer
Predicted	(+ve) Has Cancer	TP	FP
	Do not have cancer (-ve)	FN	TN

If we have More than Two classes.

Actual

Question →

Person watched Chakde?

		Actual		
		Chakde	KGF	DDLJ
Predicted	Chakde	TP	FP	FP
	KGF	FN	TN	TN
	DDLJ	FN	TN	TN

If person watched Chakde +ve

Else -ve

Predicted that these people have not watched Chakde so N

Person has not watched Chakde & Predicted also not watching chakde.

Question → Watched Chakde? Yes +ve / No -ve.

For 4 classes

		Actual			
		Chakde	KGF	DDLJ	Gadar
Predicted	Chakde	TP	FP	FP	FP
	KGF	FN	TN	TN	TN
	DDLJ	FN	TN	TN	TN
	Gadar	FN	TN	TN	TN

Chakde Nahi dekha phir bhi predicted Dekha

Chakde dekha But predicted Nahi dekha

Chakde Nahi Dekha And predicted Nahi Dekha.

*

TP (True +ve) = Correctly Identified +ve (+ve Hai → +ve Predicted)

FP (False +ve) = Incorrectly identified +ve (+ve Nahi Hai → +ve Predicted)

TN (True -ve) = Correctly Identified -ve ⇒ (-ve Hai → -ve Predicted)

FN (False -ve) ⇒ Incorrectly Identified -ve ⇒ (-ve Nahi Hai → -ve Predicted)

* 1) Accuracy = $\frac{\text{Total Correct Prediction}}{\text{Total Prediction}} = \frac{TP + TN}{TP + TN + FP + FN}$

2) Error Rate = 1 − Accuracy.

3) True +ve Rate (TPR) [Sensitivity / Recall] = $\frac{\text{Correctly Identified +ve}}{\text{Total Actual +ve}} = \frac{TP}{TP + FN}$

(Aisa +ve Jisko −ve predict kiya)

4) False Negative Rate (FNR). ⇒ $\frac{\text{Incorrectly Identified +ve}}{\text{Total Actual +ve}} = \frac{FN}{TP+FN}$ ∥

5) True Negative Rate TNR [Specificity] = $\frac{\text{Correctly Identified −ve}}{\text{Total Actual −ve}} = \frac{TN}{TN + FP}$

6) False Negative Rate [FNR] = $\frac{\text{Incorrect Identified −ve}}{\text{Total Actual −ve}} = \frac{FP}{TN + FP}$

$$\left[\quad TPR = \frac{\text{Correct +ve}}{\text{Total +ve}} \qquad TNR = \frac{\text{Correct −ve}}{\text{Total −ve}} \quad\right]$$

$$\left[\quad FNR = \frac{\text{Incorrect +ve}}{\text{Total +ve}} \qquad FPR = \frac{\text{Incorrect −ve}}{\text{Total −ve}} \quad\right]$$

Joh Joh Joh Joh

Note To Identify Heart Disease

Sensitivity = TPR = $\frac{\text{Correct Identified +ve}}{\text{Total +ve}}$ = What percentage of patient with Heart disease are Correctly

Sensitivity = TPR = Total +ve   patient with ... disease are Correctly Identified

Specificity = TNR = $\frac{\text{Correct Identified -ve}}{\text{Total -ve}}$ = What percentage of patients without heart disease are correctly Identified.

$$\boxed{FPR = 1 - \text{Specificity}} = 1 - \frac{TN}{TN+FP} = \frac{\cancel{TN} + FP - \cancel{TN}}{TN+FP} = \frac{FP}{TN+FP}.$$

Training Error →

It is prediction Error that we get when we apply the model on same data from where it is trained.

$$\mathcal{E}_{Train} = \frac{1}{n} \sum_{i=1}^{n} \text{Error}\left(f_\theta(x_i), y_i\right)$$

(for all the sample) — $f_\theta(x_i)$: Prediction of x_i — y_i: actual value.

(2) **Test Error**: It is prediction Error we get when we apply model on altogether different data set (Test set) and not on the data on which it is trained.

$$\mathcal{E}_{test} = \frac{1}{n} \sum_{i=1}^{n} \text{Error}\left(f_\theta(x_i), y_i\right)$$

for Test set //

(3) **Generalization Error** → also known as Out of Sample Error

→ Measure of how accurately an algorithm is able to predict outcome values for previously unseen data.

→ We want to know how the model will perform

→ We want to know how the model will perform on future data (we do not have today)

→ For Future we do not have X_i (input)
y_i (output)

$$\mathcal{E}_{gen} = \underbrace{\int}_{\text{over all possible value of } x \,\&\, y} \text{Error}\Big(\underbrace{f_\theta(X_i)}_{\text{Predicted}}, \underbrace{y_i}_{\text{actual}}\Big) \underbrace{P(y, x)}_{\text{How often we expect such } x \,\&\, y} \, dx .$$

Usually

$$\underline{\mathcal{E}_{train} \leq \mathcal{E}_{gen}}$$

as we do not have value of future $P(y, x)$
So we do not compute generalizⁿ Error,
we approximate it with <u>Testing Error</u>.