** Important Keys for basic ML **

① **Train data** – We design our de model to predict. The train data that is use to train an algorithm or machine learning model to predict.

② **Test data** – The test dataset is a subset of train dataset that is utilized to give an objective evalution of a final model.

⇒ In the context of the validation dataset, assesment of model skill on unknown data.

③ **Train input** – The train input data typically refers to the input data used for traing a model. This data consist of a matrix or table, where each row represent an individual data sample, and each column correspond to a specific feature or attributs.

* **Train output** – Train output refers to the corrospondi target or dependent variable in supervised, learning task.

It represent the expected or known outcomes or labels associated with the input data used during the model traing phase.

④ **Test input** – Test input refers to the unseen data used to evaluted a trained models performance, serving as

The set of features feed into the model for prediction or classification. The test input is used for testing a regression, performance.

* Test output — Test output is the predicted or expected outcome generated by a trained model when its given unseen input data during the evalution phase.

(5) = Which data is used to develop the model?

Ans= Train data is the data that is used to develop the model to accurately predict a particular outcome.

(6) = Which data is used to test the model?

Ans= The model is built (with our traing data), we need unseen data to test model. This data is called testing data, and we can use it to evaluate the performance and progress of our algorithms, traing and adjust or optimize it for improved results.

(7) = Which data will compairs that gives test error?

Ans= we pass test input data to the model, and we compare these prediction with test output.

(8) = Whic data will compare that gives train error?

Ans= We will pass the train input data to the model and we compare these prediction with train output.

(9) - What is underfit ?

Ans = Underfitting means the model is too simple to capture the patterns in the traing data, resulting in poor performances and mobility to accurately predict.

⇒ Train error is high, high bias - low variance

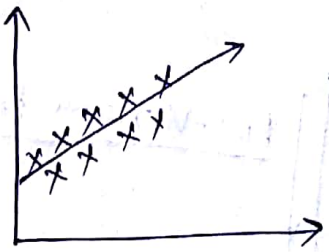
Underfit

(10) - What is normal fit ?

Ans = Normal fit refer to a statistical process used to estimate the mean and standard deviation of data, making it look like a normal distribution. allowing better understanding and analysis of dataset.

⇒ Both train error and test error are low.

⇒ low bias - low variance



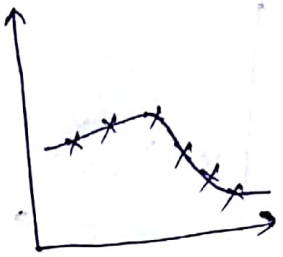
Normal fit

(11) - Overfit ?

Ans= Overfitting is happened when a model gives accurate predictions for training data but not for new data.

⇒ Train error is low and test error is high.

⇒ low bias and high variance.



overfit

*(12) - What is bias-variance curve?

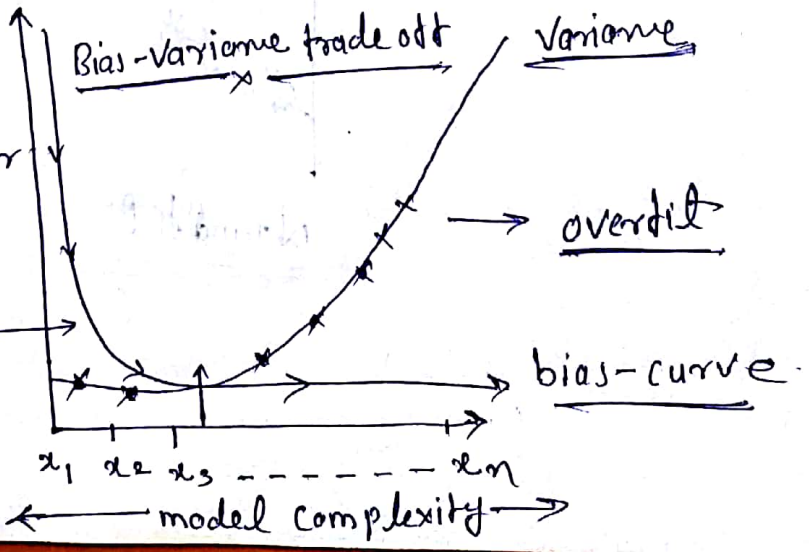
Ans= The bias-variance trade off represent the balance between a model simplicity (bias and its ability to adopt to different data points (variance) to minimize errors in predictions.

⇒ Bias curve will explain about train error.

⇒ Variance curve will explain about the test errors.

⇒ Variance ⇒ test error

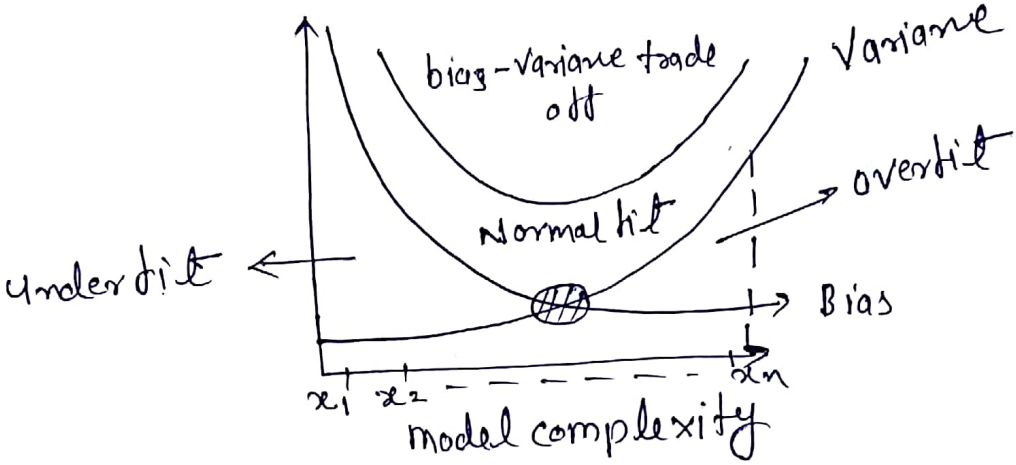
⇒ bias curve → train error



(13) - What is the mean by model complexity?

Ans= model complexity refers to how a model is capturing relationships within data, with higher complexity model being more detailed and capable of learning intricate pattern but being prone to overfitting.

⇒ It means how flexible as intricate the model structure.



(14) - What happen it add so many variables?

Ans- Model complexity is increasing as we increase number of variance.

(15) - What is low bias- low variance?

Ans= Normal fit.

(16) - What is low bias-high variance?

Ans= overfit.

(17) - What is high bias- low variance?

Ans= Underfit.

Avinash..