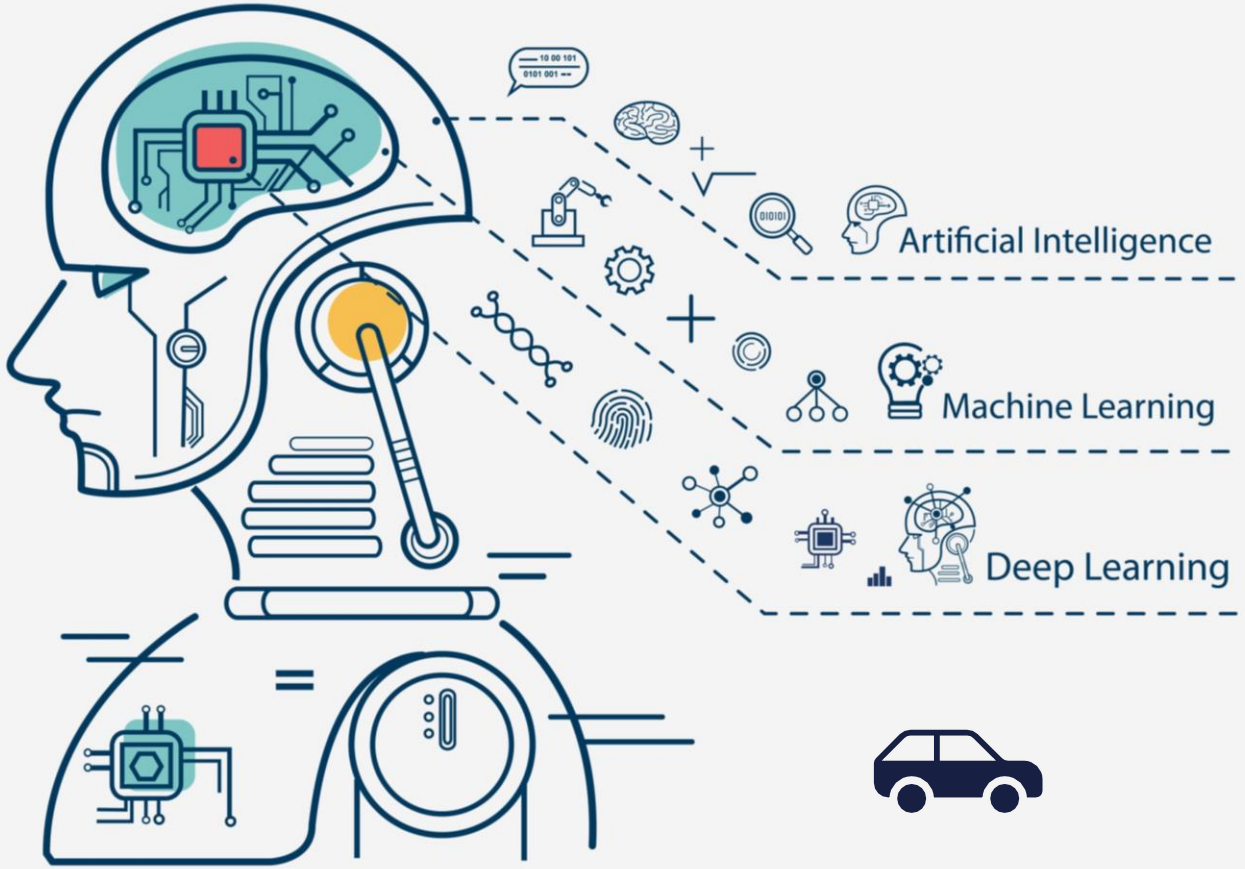
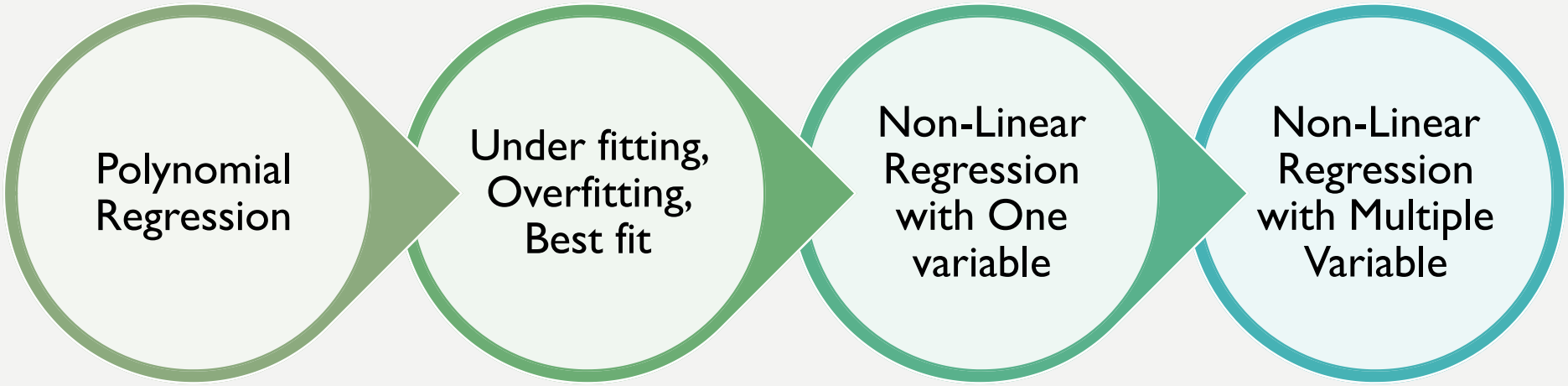
APSSDC

Andhra Pradesh State Skill Development Corporation



MACHINE LEARNING USING PYTHON

DAY3 AGENDA



By Anil Kumar APSSDC

POLYNOMIAL REGRESSION

NON LINEAR REGRESSION

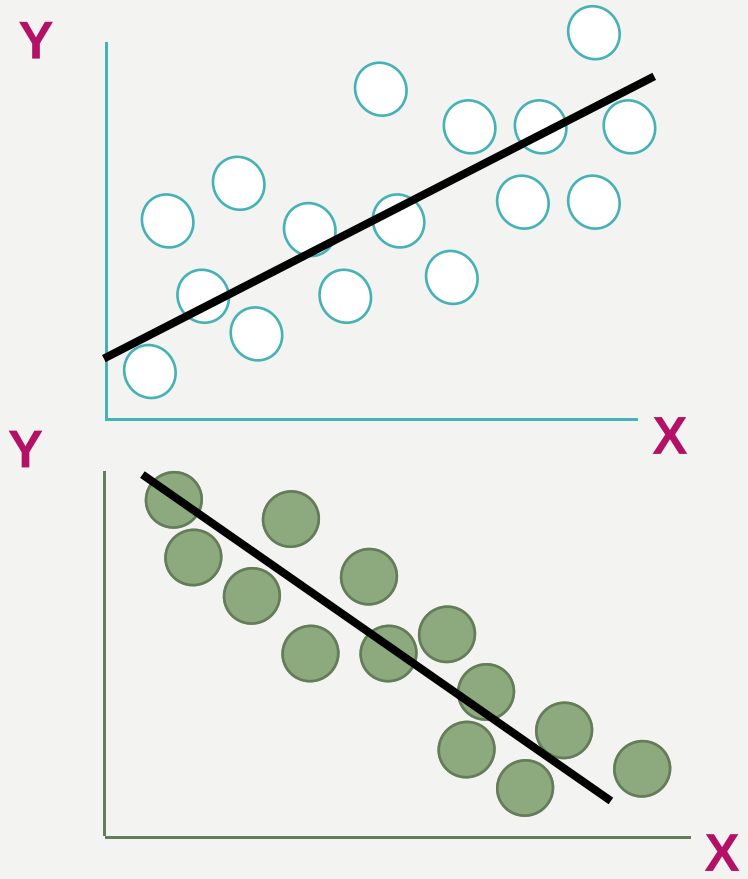
It is a special type of multiple regression whose independent variables are powers of a single variable X. It is used to approximate a curve with unknown functional form.

$$Y_i = \alpha + \beta_1 X + \beta_2 X^2 + \ldots + \beta_k X^k + \varepsilon_i$$

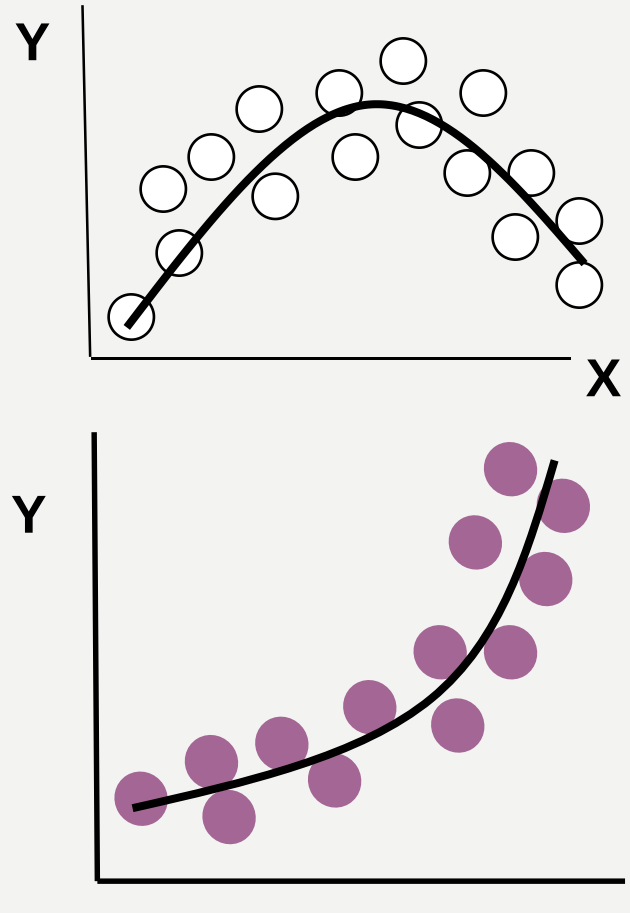
By Anil Kumar APSSDC

LINEAR CORRELATION

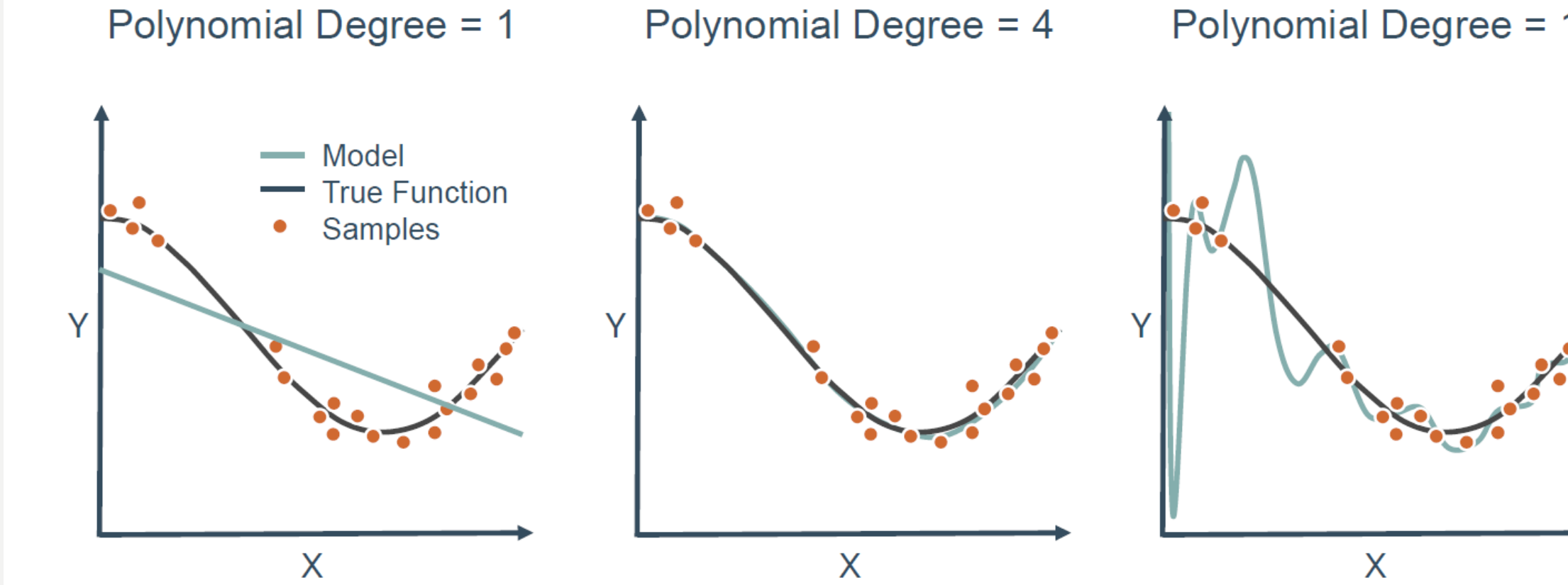
Linear relationships



Curvilinear relationships



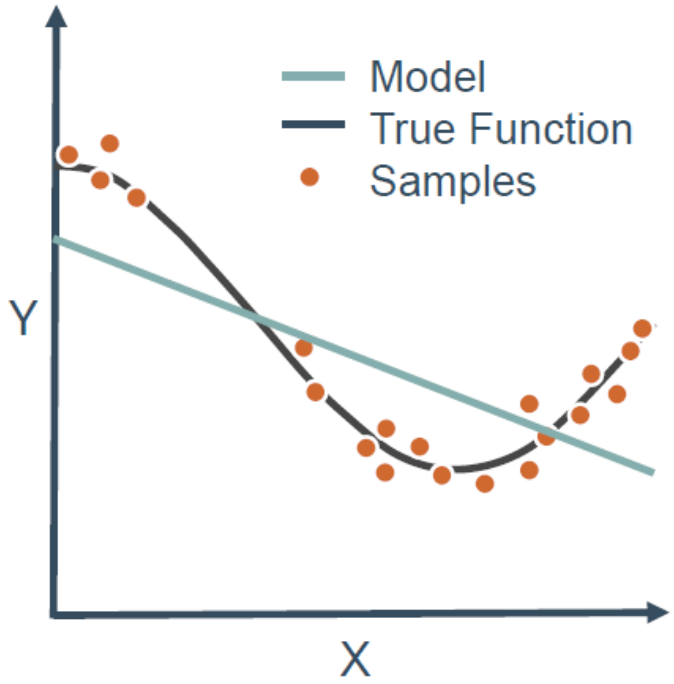
CHOOSING BETWEEN DIFFERENT COMPLEXITIES



By Anil Kumar APSSDC

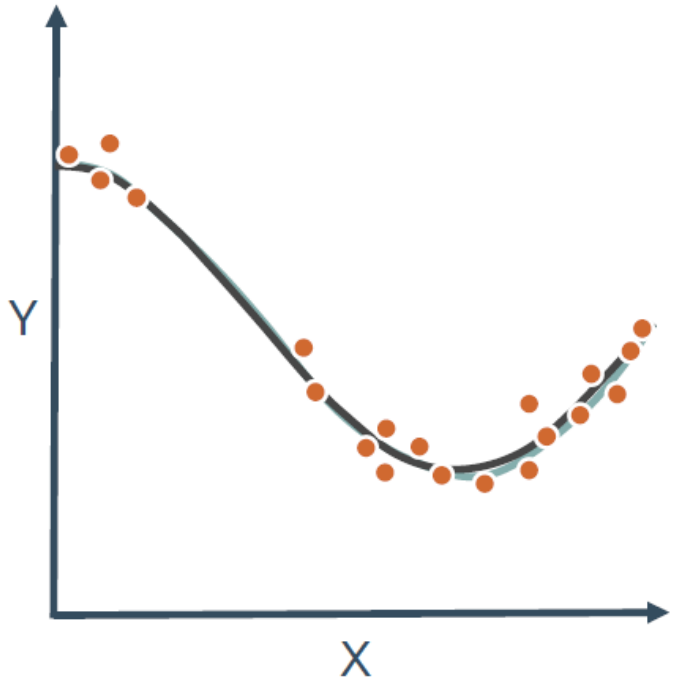
BIAS–VARIANCE TRADEOFF

Polynomial Degree = 1



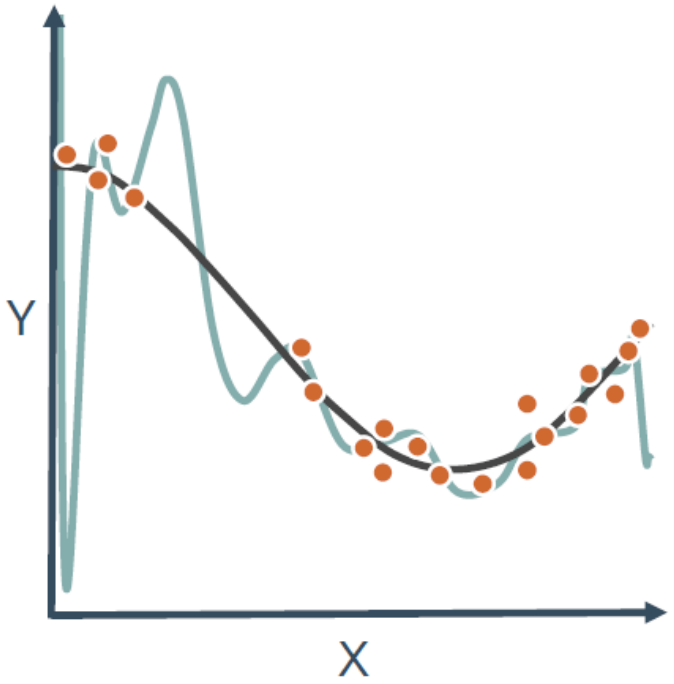
High Bias
Low Variance

Polynomial Degree = 4



Just Right

Polynomial Degree = 15



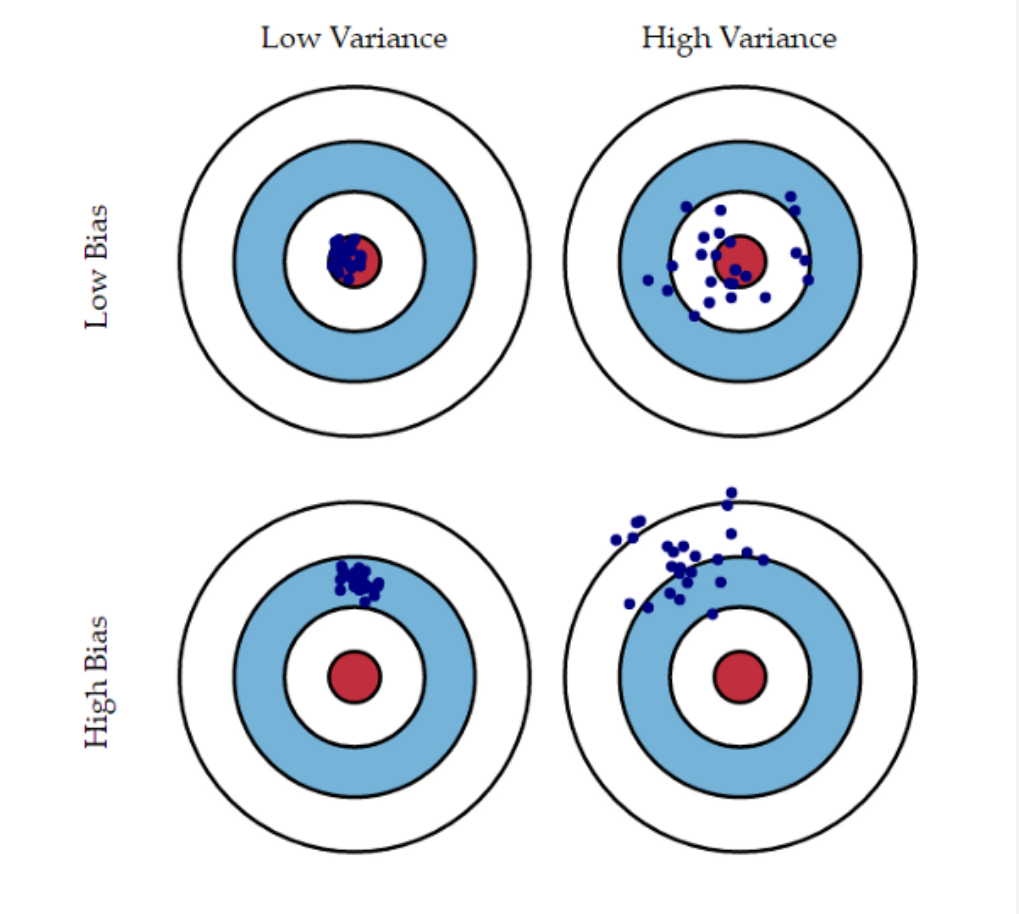
Low Bias
High Variance

By Anil Kumar APSSDC

What is bias and variance?

Bias is the difference between the average prediction of our model and the correct value which we are trying to predict. Model with high bias pays very little attention to the training data and oversimplifies the model. It always leads to high error on training and test data.

Variance is the variability of model prediction for a given data point or a value which tells us spread of our data. Model with high variance pays a lot of attention to training data and does not generalize on the data which it hasn't seen before. As a result, such models perform very well on training data but has high error rates on test data.



By Anil Kumar APSSDC

UNDERFITTING

In supervised learning, **underfitting** happens when a model unable to capture the underlying pattern of the data. These models usually have high bias and low variance. It happens when we have very less amount of data to build an accurate model or when we try to build a linear model with a nonlinear data.

By Anil Kumar APSSDC

OVER FITTING VS UNDER FITTING

- Under Fitting

- Over Fitting

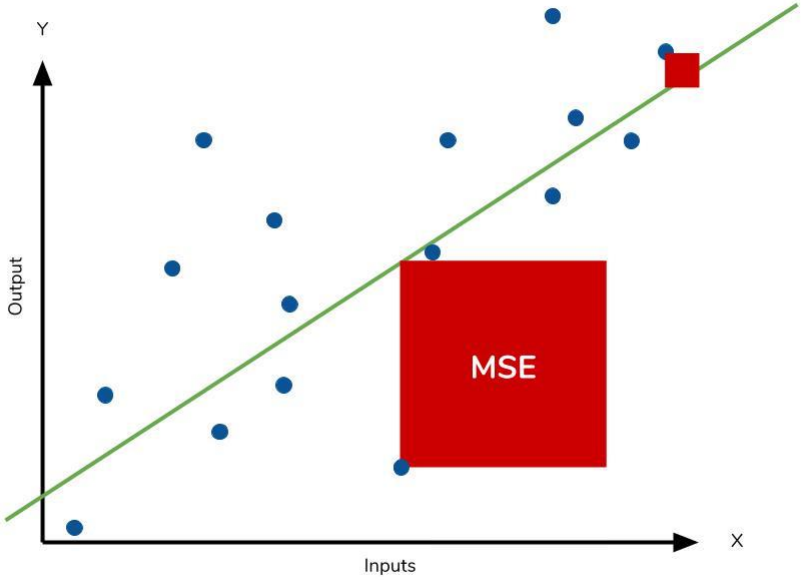
By Anil Kumar APSSDC

Contd..

2. Mean Square Error

$$MSE = \frac{1}{n}\sum\underbrace{\left(y - \hat{y}\right)}^{2}$$

The square of the difference between actual and predicted



Evaluation Metrics

1. Mean Absolute Error

$$MAE = \frac{1}{n}\sum \left| y - \hat{y} \right|$$

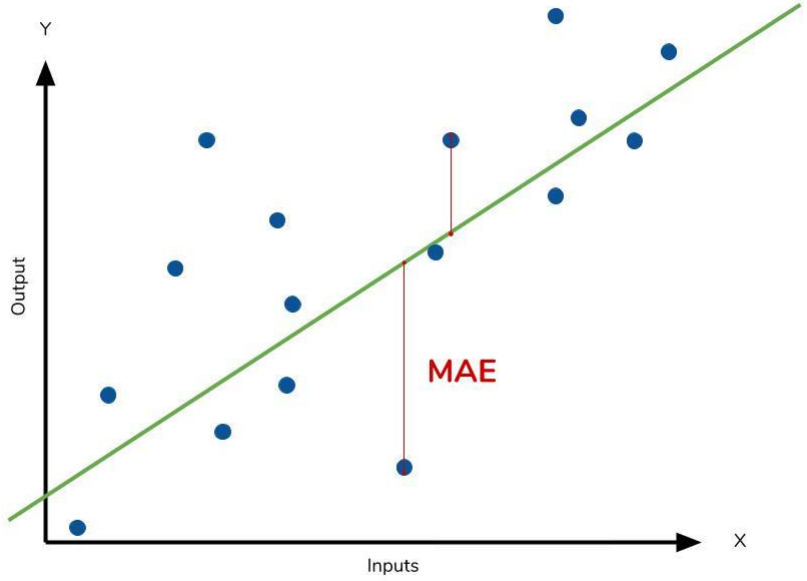
Divide by the total number of data points

Predicted output value

Actual output value

Sum of

The absolute value of the residual



Contd..

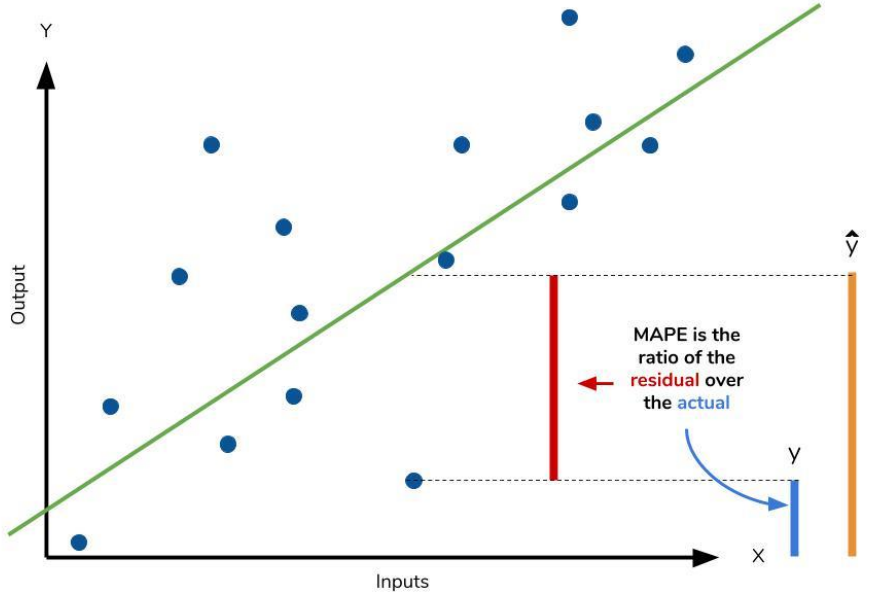
3. Mean Absolute Percentage Error

$$MAPE = \frac{100\%}{n} \sum \left| \frac{y - \hat{y}}{y} \right|$$

Multiplying by 100% converts to percentage

The residual

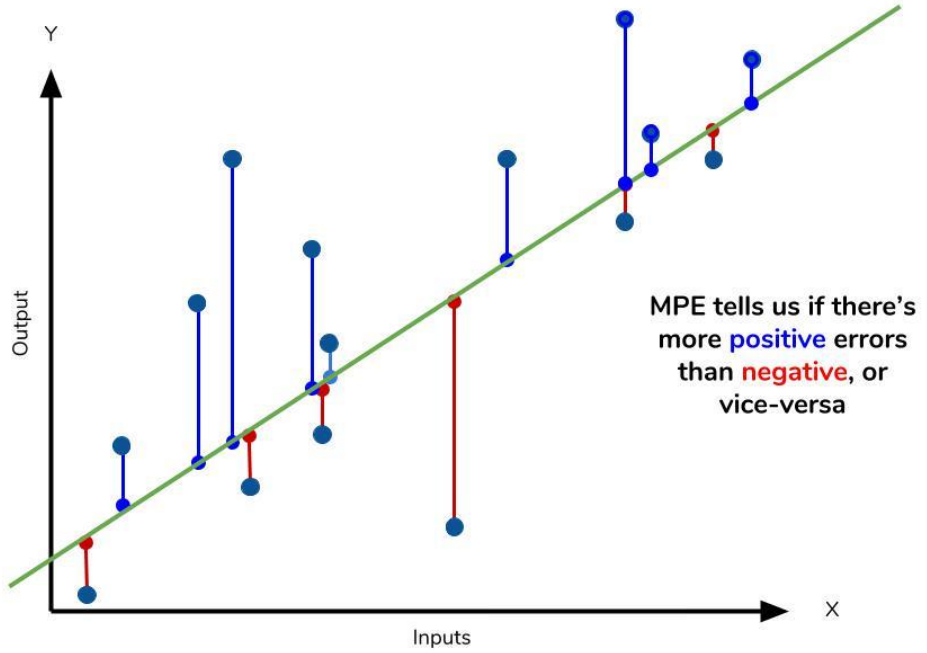
Each residual is scaled against the actual value



Contd..

3. Mean Percentage Error

$$MPE = \frac{100\%}{n} \Sigma \left(\frac{y - \hat{y}}{y} \right)$$



Conclusion

Acroynm	Full Name	Residual Operation?	Robust To Outliers?
MAE	Mean Absolute Error	Absolute Value	Yes
MSE	Mean Squared Error	Square	No
RMSE	Root Mean Squared Error	Square	No
MAPE	Mean Absolute Percentage Error	Absolute Value	Yes
MPE	Mean Percentage Error	N/A	Yes

https://www.dataquest.io/blog/understanding-regression-error-metrics/