Polynomial regression
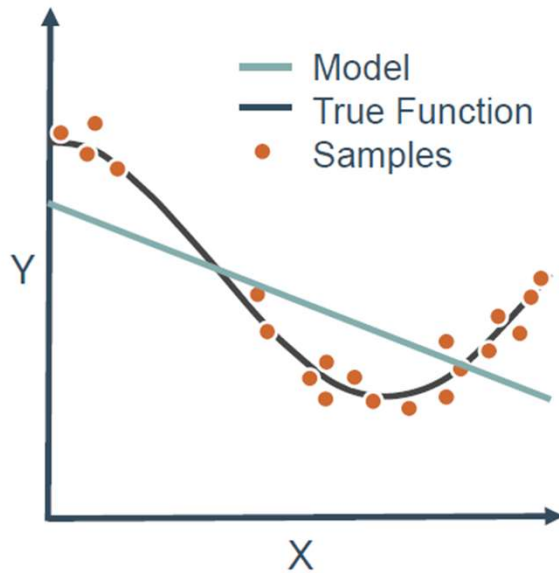
- It is a special type of multiple regression whose independent variables are powers of a single variable X. It is used to approximate a curve with unknown functional form.
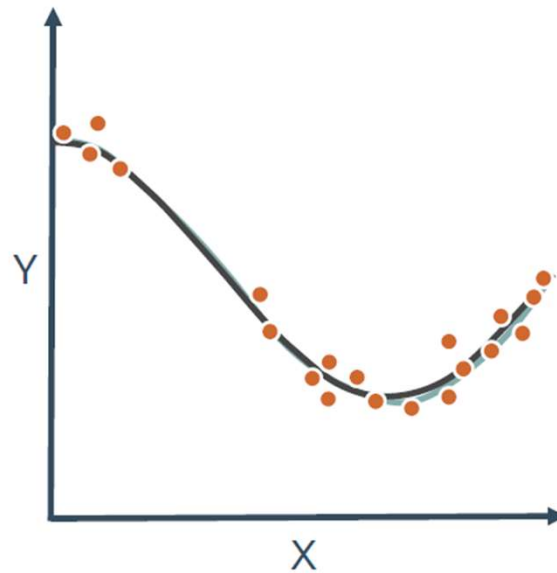
$$Y_i = \alpha + \beta_1 X + \beta_2 X^2 + \ldots + \beta_k X^k + \varepsilon_i$$
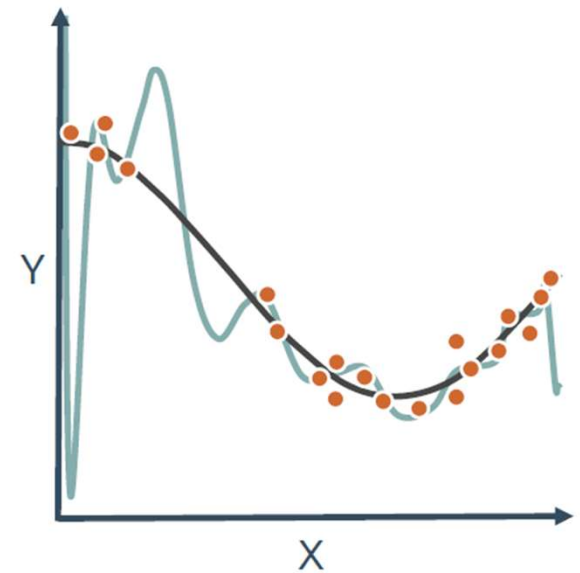
CHOOSING BETWEEN DIFFERENT COMPLEXITIES
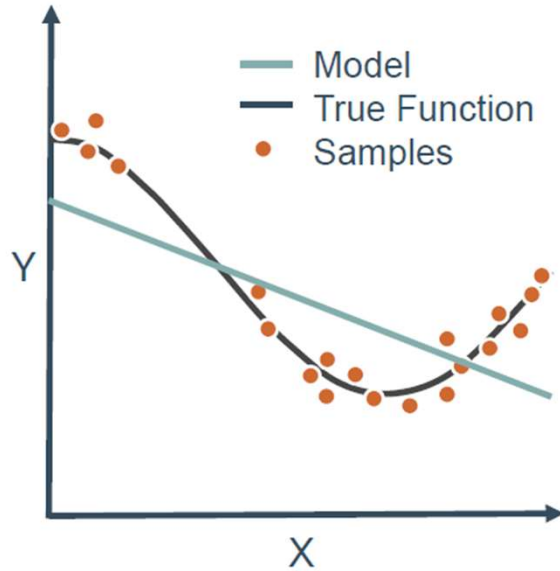

Polynomial Degree = 1   Polynomial Degree = 4   Polynomial Degree = 15

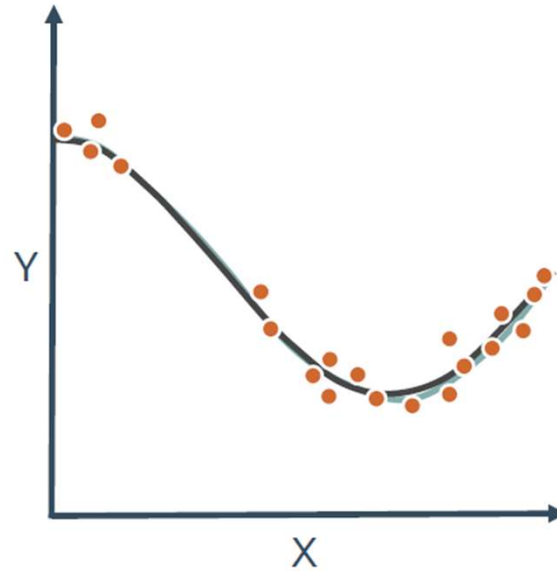Legend: Model, True Function, Samples

BIAS—VARIANCE TRADEOFF

Polynomial Degree = 1

Polynomial Degree = 4

Polynomial Degree = 15

— Model
— True Function
● Samples

Y

X

Y

X

Y

X

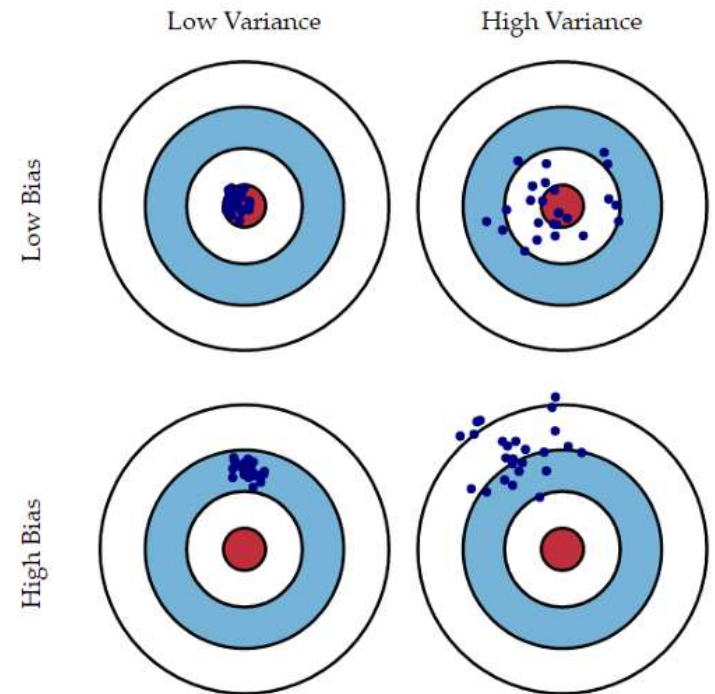**High Bias
Low Variance**

Just Right
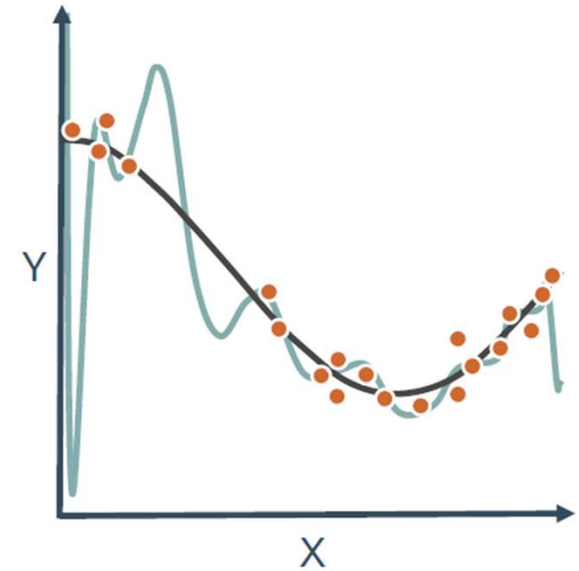
**Low Bias
High Variance**

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.

Low Variance   High Variance

Low Bias

High Bias

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.

Over Fitting vs Under Fitting

Under Fitting   **Over Fitting**