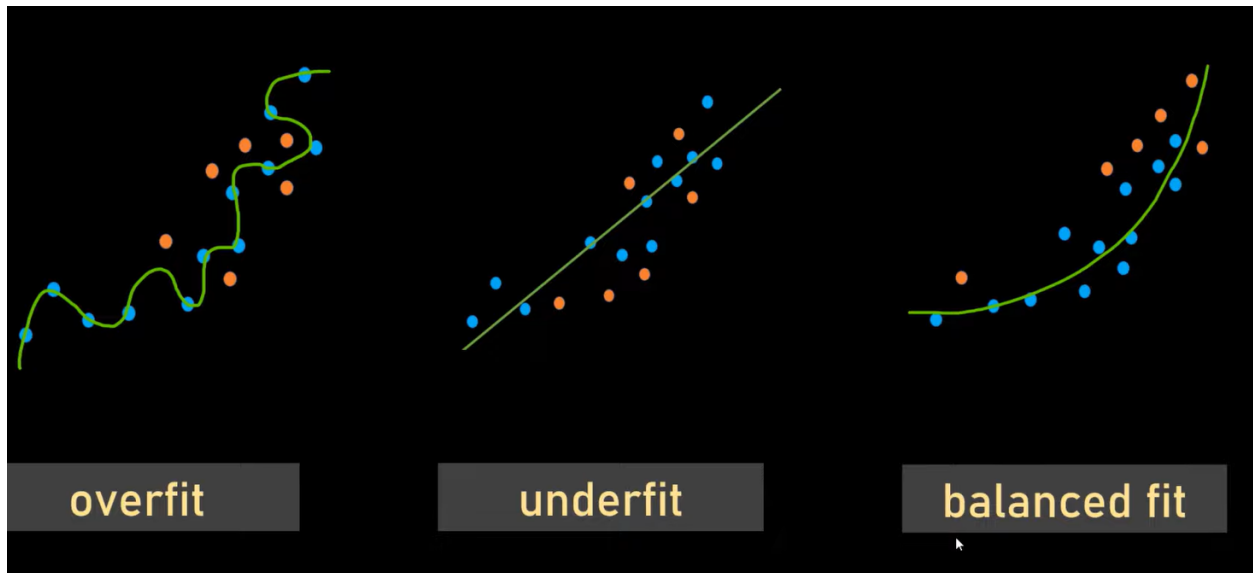


# Bias vs Variance

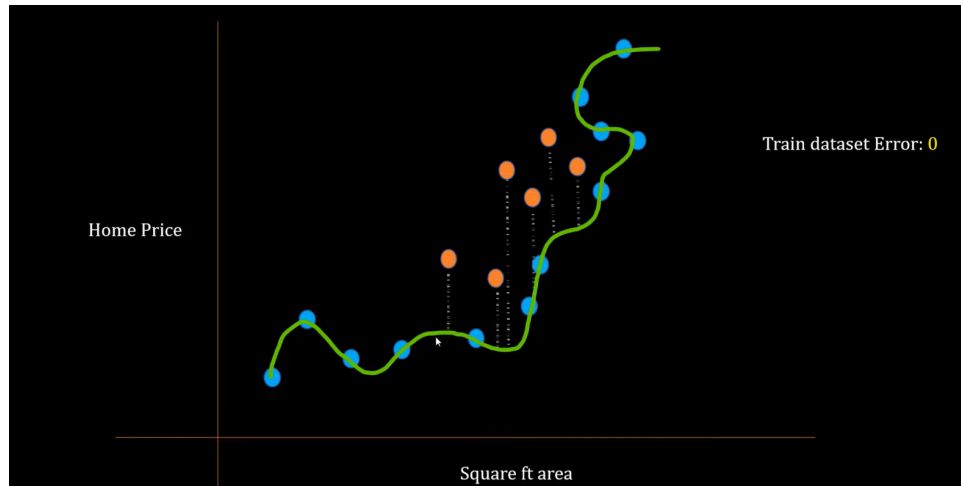
Ahman Smith



There are three classifiers:

- Overfit (conforms too much to the dataset)
- Underfit (conforms too little to the dataset)
- Balanced fit (just right)

An overfit model tries to fit exactly to the training sample where the error becomes close to 0, **but the test error increases.**

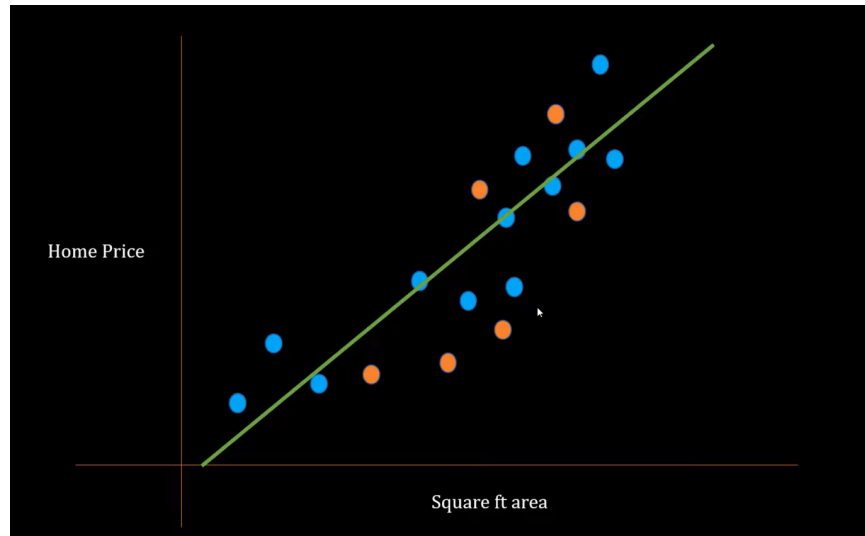


The orange dots are the test data. When it conforms directly to the blue dots, the error increases a lot.

### Important

- Test data errors greatly based on your selection of data points
- In the figure above, it's an example of high variance
- Overfit models have **high variance** (when your test error varies greatly on the datapoints you select)

## Underfitting



- Training error is actually high, and your test data is also high
- When you select a different set of data points, the line will be different, but the error will be similar.
- **An example of low variance**
- Our test error doesn't vary that much based on our dataset

#### **Bias in Machine Learning:**

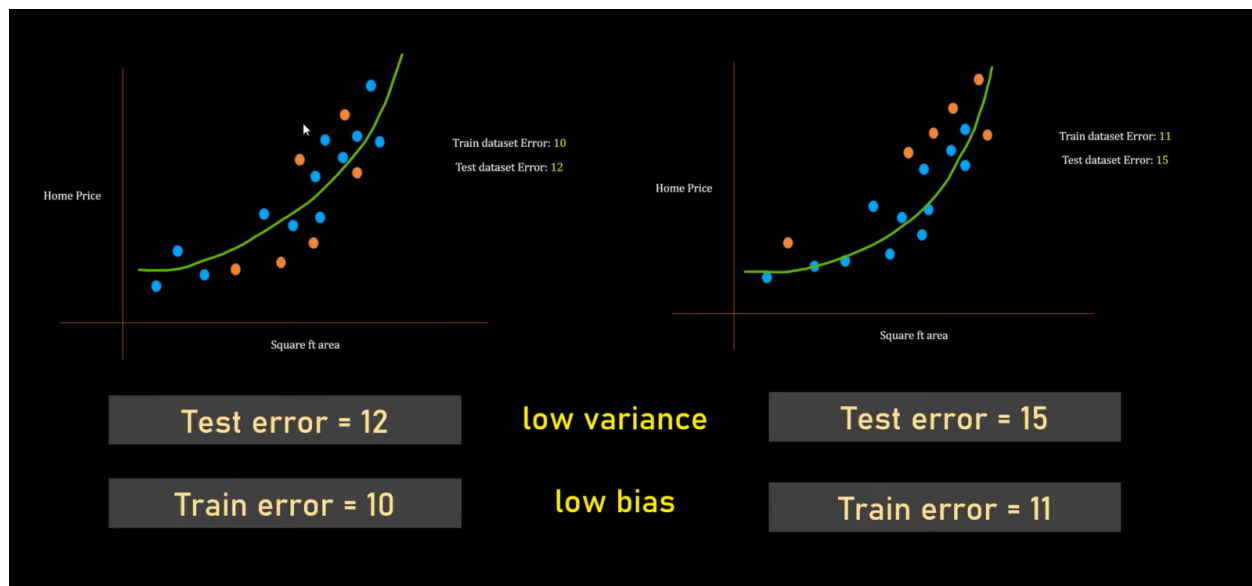
- Train error is high

High bias — a measurement of how accurately a model can capture a pattern

#### **Variance:**

- Test error is high

## **Low Bias, Low variance**



- Train error and test error are both low
- Model selection is such that both errors are low no matter the dataset
- The test error doesn't vary too much depending on the training samples, the train error is low in general

**Closer to truth = low bias**

**Far away from truth = high bias**

**Data is super scattered: high variance**

**High**

**WORST CASE:**

- High bias
- High variance
- Far from truth