Bias is the error that calculates the difference between the average prediction of our model and the actual value that we are trying to predict.

A model suffering from high bias is a simple model which pays very little attention to the training data. This type of model always leads to a high error on both **training and test data**. Let's take an example. Suppose we want our model to predict the animal by showing photos of animals. We trained the model on only one attribute **poiniting_ears.** Then we showed the image of a cat to the model. **So the model predicted it as a fox also has pointed ears.**

Actual   **Predicted**

This shows the model is not able to capture other details while predicting as it has bias.

Characteristics of a high bias model include:

- Not able to capture proper data trends

- Trained over noise also. So giving less accurate results

- Suffers from underfitting

- A more general or simple model

What is variance?

Variance is the opposite of Bias. Variance is also an error that measures the randomness of the predicted value from the actual value.

Variance can be defined as the model's sensitivity to fluctuations in the data. if we model is allowed to view the data too many times, it will learn very well for only that data. It will capture most patterns in the data, but it will also learn from the unnecessary data present, or from the noise. When we train our model with too much data or allow it to view the data too many times, it will learn the data including noise, which will cause our model to consider trivial features as important. In this case, our model is overfitted. Now let's continue the above example of **animal prediction**. If we consider **fur** as a feature then that will be noise as many animals have fur.

Note: *Noise* here means *irrelevant details* which are not required for the predicting output.

If you will train the model with some 100 images of cat and dog and again show the same images to it. It will predict correctly. But if you will some different cat and dog images the model will not be able to predict it correctly. This model performs well during the training phase but not during a test phase. And it might be looking at specific features like the nose and ears also. When the variance is high our model will capture all the features of the data given to it will tune itself to the data and predict it very well.

A model should have less variation in the predicted values with changes in the training data set. *Continuing the same cat example,* now this time we gave more features to the model for training

Variance errors are either the **low variance or high variance.**

- *Low variance: A model has a small variation* in the predicted values with changes in the training data set.
- *High variance: A model has a high variation* in the predicted values with changes in the training data set. A model having high variance learns everything shown to it and performs well with the training dataset, but not on test data.

Understanding with example

Suppose you want to predict house price with respect to the house area.

Let's say all these **blue dots** are **training samples**, the **orange dots** are **test samples** as shown in the figure below. We can train a model that fits these blue dots perfectly which means our model is an overfitted model. An overfit model tries to fit exactly to the training samples but not to the test samples that's why training error becomes close to zero and test error is high.

Calculating error

Nonlinear model

Now let's say you want to figure out an error for this particular orange test data point. The error will be this gray dotted line. And you can measure the error for all your test data set and average it out.

Let's say you get this error as 100(as shown in the left figure below). When you split your dataset, you pick your training samples at random. Suppose your friend, uses the same model, the same methodology but might be choosing a different set of training samples. In both scenarios training data set error will be zero because you both are trying to overfit the model. Let's say you get this test error as 100(as shown in the first figure) and your friend gets a test error of 27(as shown in the second figure). *Why you are getting high errors as compared to your friend* even after using the same methodology and same data.

This is because the test error varies greatly based on your selection of training data points. And this is called high variance because there is high variability in the test error based on what kind of training samples you are selecting. Now you are selecting training samples at random, so your test error varies randomly which is not good and this is the common issue with overfit model.

Now next question that comes to mind is what if we are using linear models.

Linear model

When you select a different set of training data points, your training and test data set error is still kind of similar. This means variability is not there much.
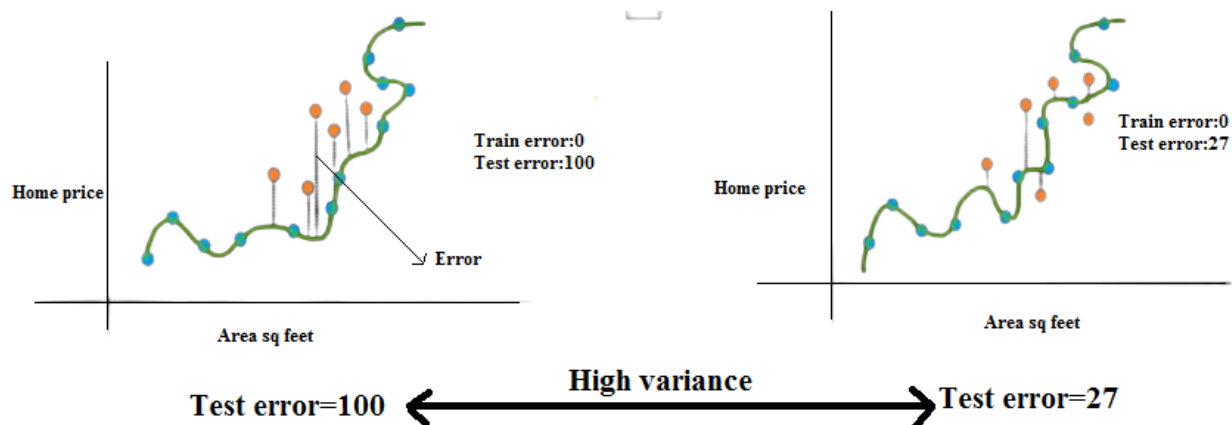
Examples of bias and variance

Some machine learning algorithms with low bias **are k-Nearest Neighbours, Decision Trees, and Support Vector Machines**. At the same time, some machine learning algorithms that have high bias are **Linear Regression and Logistic Regression.**

Summary

Bias——>Underfitting—->High train and test error

Variance—->Overfitting——>High test error

Bias variance tradeoff

Till now we got the idea that in order to avoid overfitting and underfitting in the model we have to decrease bias and variance.

If the model is having fewer parameters, it may have low variance and high bias. Whereas, if the model is complex with a large number of parameters, it will have high variance and low bias. So, there is a need to strike a balance between bias and variance errors, and this balance between the bias error and variance error is known as the *Bias-Variance trade-off.*

Underfitting
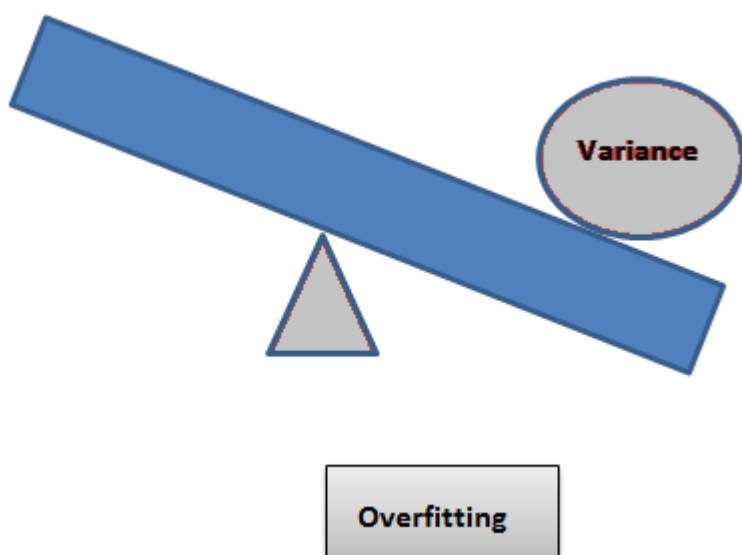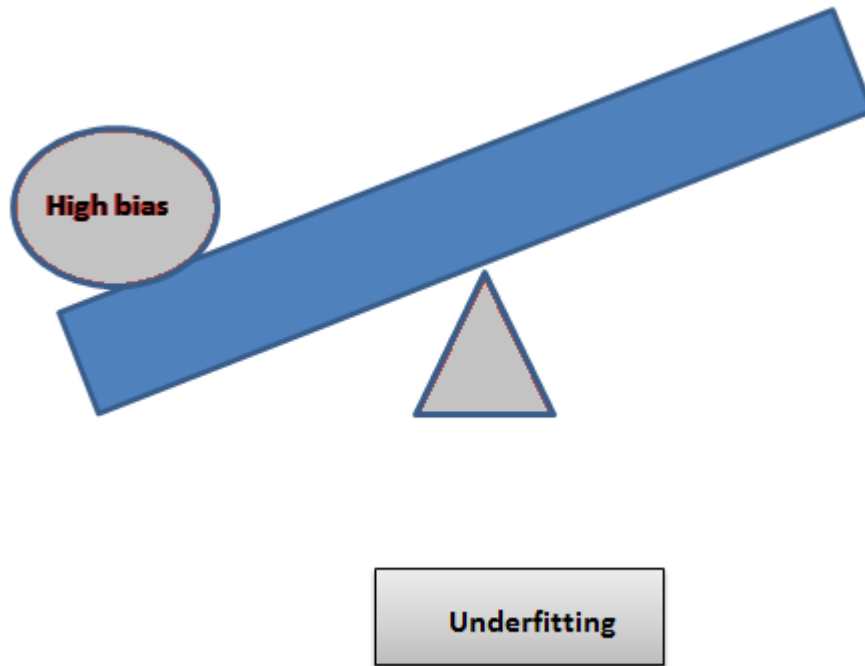
Note: When the model is suffering from high bias then that means it is suffering from low variance and vice versa.

Consider this **Bull's eyes diagram.** The center i.e. the bull's eye is the model result we want to achieve that perfectly predicts all the values correctly. As we move away from the bull's eye, our model starts to make more and more wrong predictions.
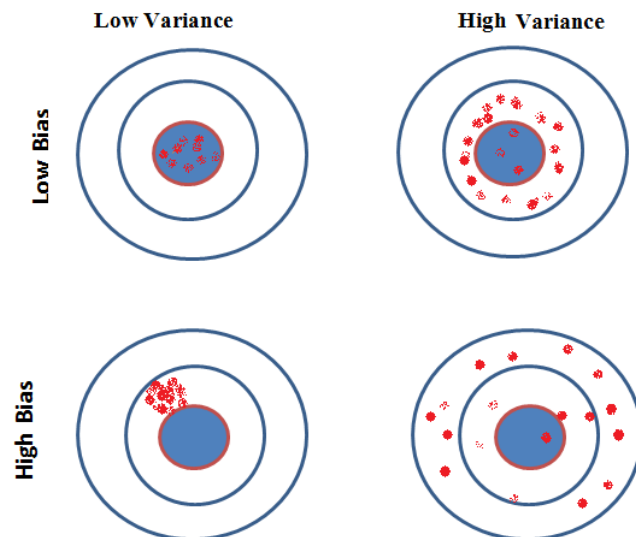
- **Low-Bias,   Low-Variance:**
  The combination is an ideal machine learning model. However, it is not possible practically.
- **Low-Bias, High-Variance:** This is a case of **overfitting** where model predictions are inconsistent and accurate on average. The predicted values will be accurate(average) but will be scattered.
- **High-Bias, Low-Variance:** This is a case of **underfitting** where predictions are consistent but inaccurate on average. The predicted values will be inaccurate but will be not scattered.

- **High-Bias, High-Variance:** With high bias and high variance, predictions are inconsistent and also inaccurate on average.