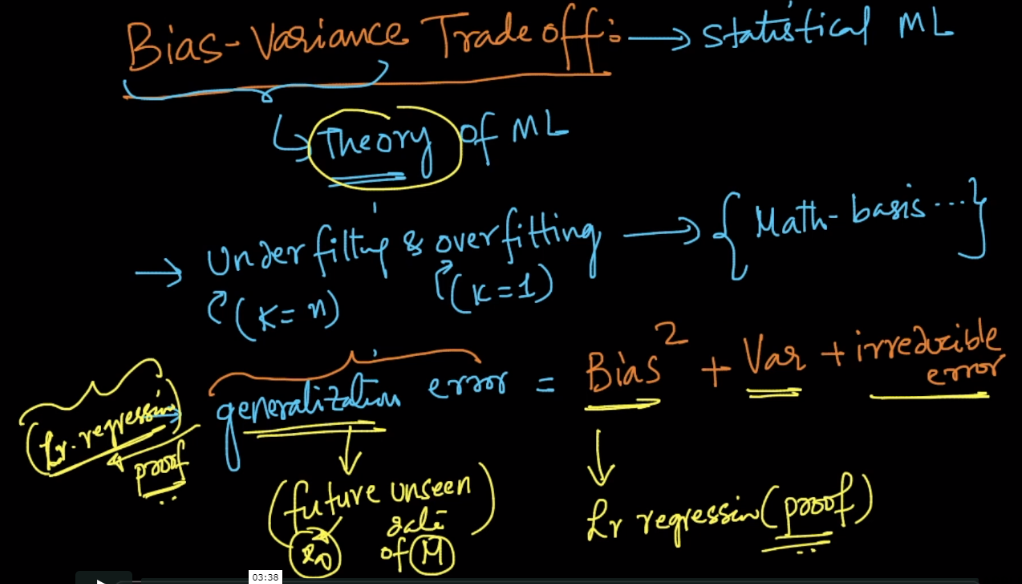
Bias-Variance Trade off is basically the Mathematical way of underfitting and overfitting.

**Generalization Error:** Error we get when model is applied test data. Or 1-accuracy when test data or future unseen data is applied on model.

Generalization error = bias2 + variance + irrudcible error.

Irrudicable errors are those which can’t be detected or removed.



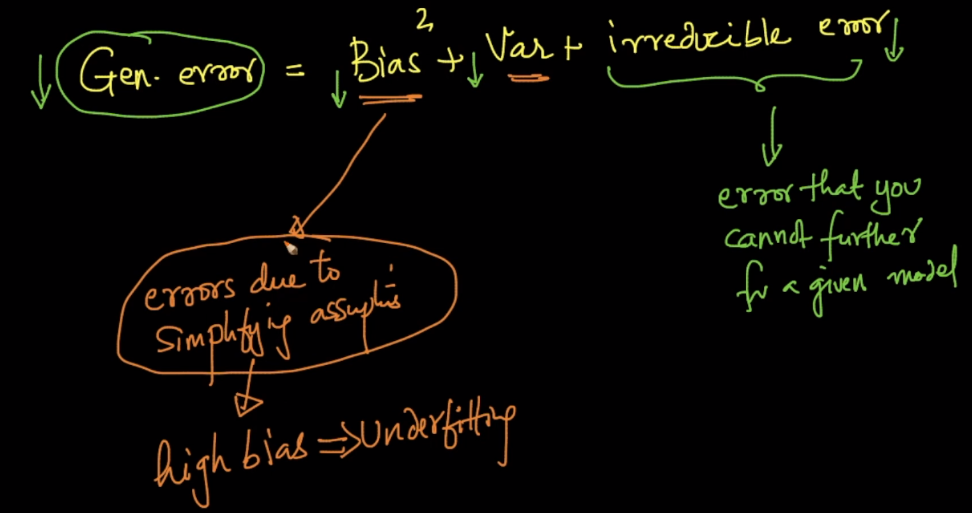
**Bias2:**

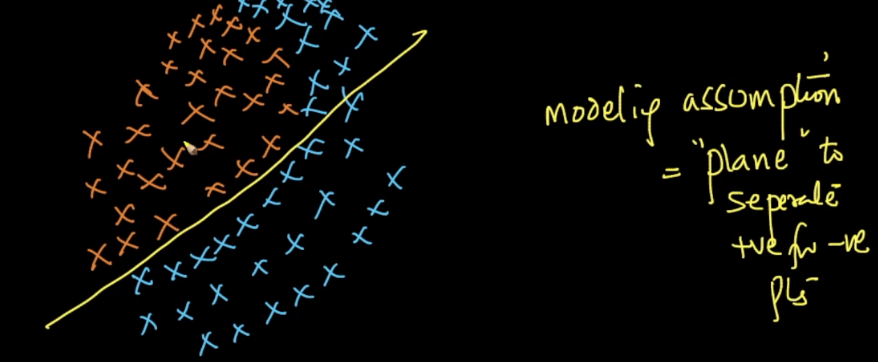
It occurs because of simplifying assumption. Example suppose our dataset is like that, it can’t be separated using line or a plane, it needs curve shape in order to separate two classes, instead of that you are providing line or a plane, that means you are simplifying the process of classification, but this will lead to wrong classification of some points.

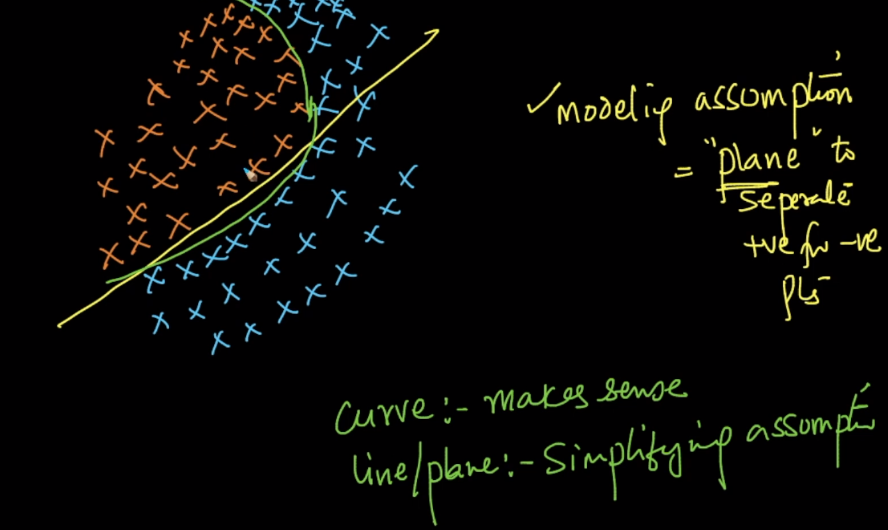
This is also termed as **Under Fitting.**

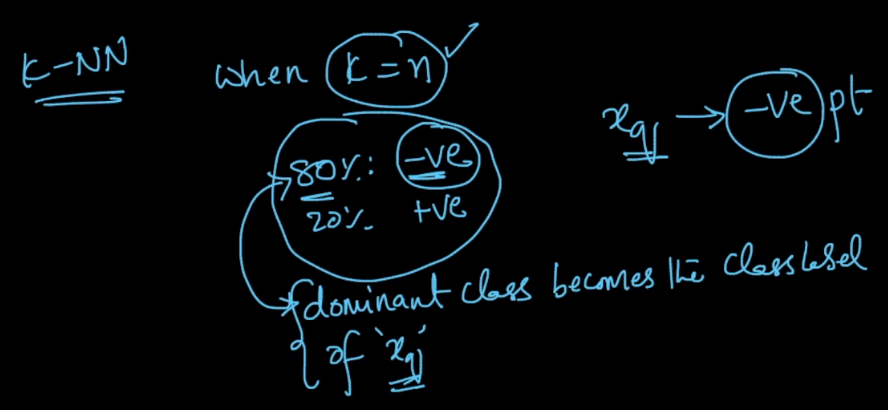
Example: Using K-NN with K=n, where n is total no. of points, there are 80% -ve, and 20% +ve.

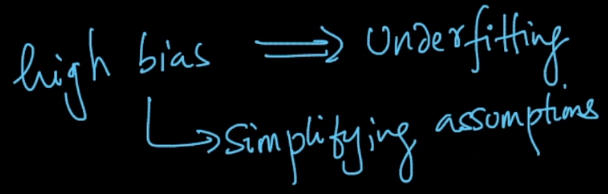
So now if any new datapoint comes up it will always predict it as -ve because of Majority voting, that means we are simplifying the model to give only one result as most of the points belongs to that result.











**Variance:**

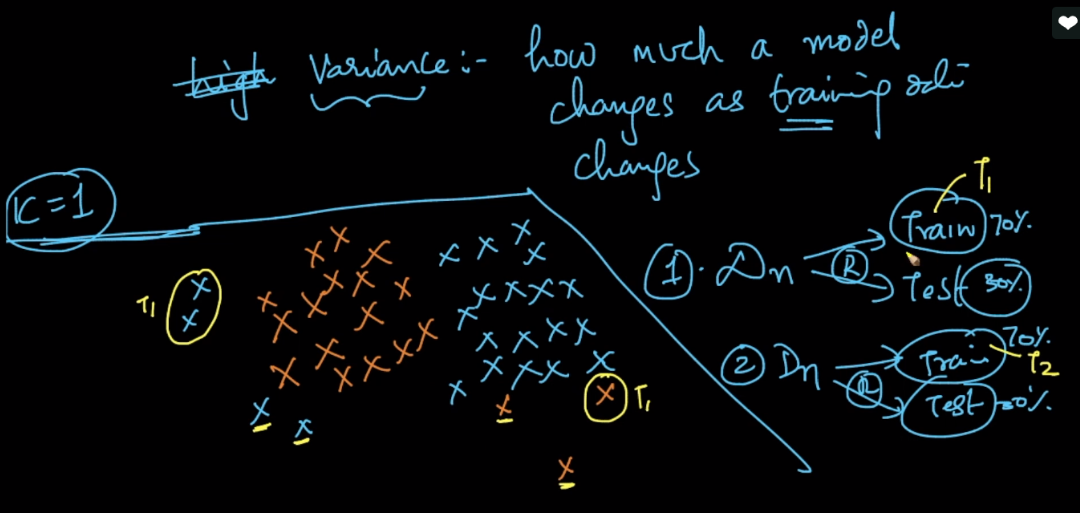
Variance tells how much a model changes (gives different result) when training data changes.

Suppose we are using k-NN with k=1.

We are randomly splitting dataset into train and test. So For first time splitting train contains all the points along with points present in marked circle and it doesn’t contains point which have underline.

And For second time splitting train data contains all data along with underlined data, but it doesn’t contain points within circle.

So here is small difference in train data for both time splitting.

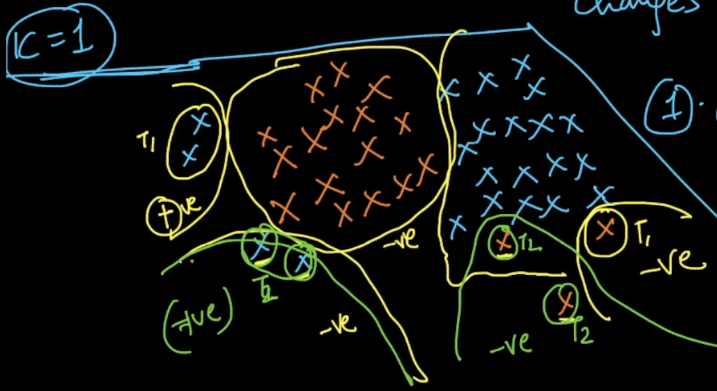


Now we train model using train data of first time splitting. Below image shows how decision boundary is created.



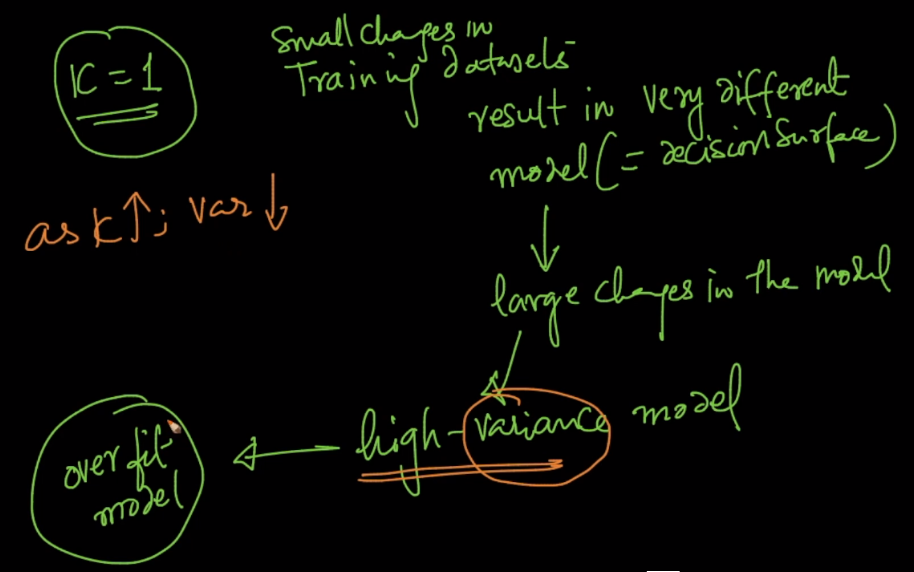
Now we are training a model using train data obtained in second time splitting, and below image shows the decision surface for this too..

We can see that since we are using same dataset for splitting, and because of small changes in train whole boundaries which were predicting -ve becomes +ve and vice versa.



So what we can see is if k increases, variance decreases, because the decision tends to become more smother.

If a model has high-variance then that termed as overfitting.



So we need to find the balance of bias of variance to get a good performance.

