VARIANCE AND BIAS:

Estimator:it is an equation for picking the best or most likely accurate,datamodel based upon the observation in reality.

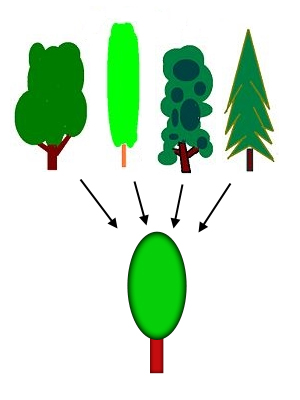
Classifier vs estimator vs model:

* an estimator is a predictor found from regression algorithm
* a classifier is a predictor found from a classification algorithm
* a model can be both an estimator or a classifier

Generalization error:

It is known as out of sample error.it says how accurately your algorithm is able to predict outcome values for previously unseen data.

Example:

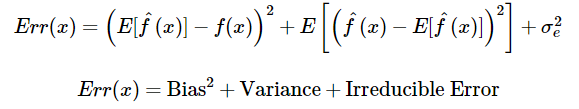


Prediction errors:(bias and variance)

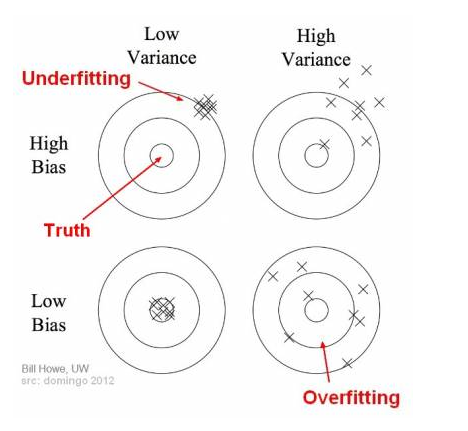
BIAS:it is nothing but the difference between avg prediction of our model and the correct value which we are trying to predict.  Model with high bias pays very little attention to the training data and oversimplifies the model. It always leads to high error on training and test data.

VARIANCE: **Variance**, in the context of **Machine Learning**, is a type of error that occurs due to a model's sensitivity to small fluctuations in the training set.

It is nothing but variability of model prediction for given datapoint. 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.



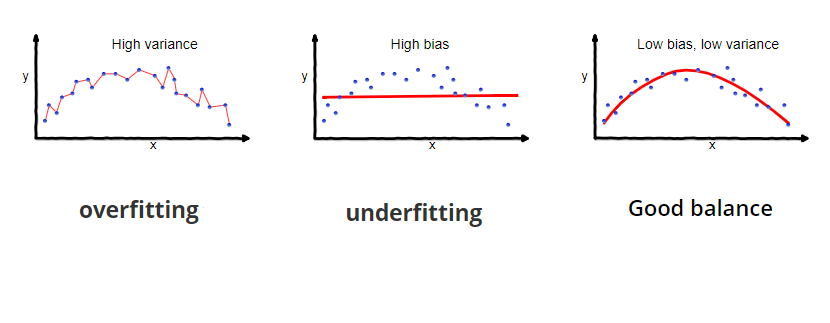
Irreducible error is the error that can’t be reduced by creating good models. It is a measure of the amount of noise in our data. Here it is important to understand that no matter how good we make our model, our data will have certain amount of noise or irreducible error that can not be removed.



In the above diagram, center of the target is a model that perfectly predicts correct values. As we move away from the bulls-eye our predictions become get worse and worse. We can repeat our process of model building to get separate hits on the target.

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. Also, these kind of models are very simple to capture the complex patterns in data like Linear and logistic regression.

In supervised learning, **overfitting** happens when our model captures the noise along with the underlying pattern in data. It happens when we train our model a lot over noisy dataset. These models have low bias and high variance. These models are very complex like Decision trees which are prone to overfitting.



BIAS AND VARIANCE TRADEOFF:

If our model is too simple and has very few parameters then it may have high bias and low variance. On the other hand if our model has large number of parameters then it’s going to have high variance and low bias. So we need to find the right/good balance without overfitting and underfitting the data.

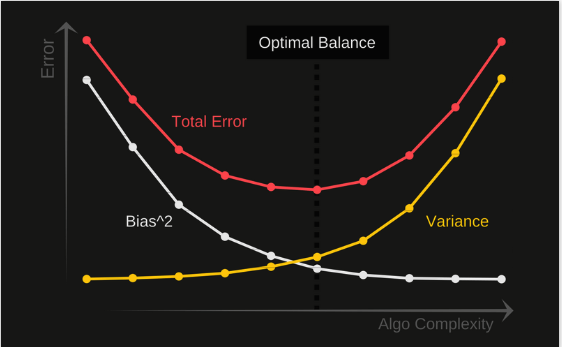
This tradeoff in complexity is why there is a tradeoff between bias and variance. An algorithm can’t be more complex and less complex at the same time.

**Total Error**

To build a good model, we need to find a good balance between bias and variance such that it minimizes the total error.

https://miro.medium.com/max/60/1*SKHGhoGKnBh_GPGHI2Ktvw.png?q=20

https://miro.medium.com/max/551/1*SKHGhoGKnBh_GPGHI2Ktvw.png

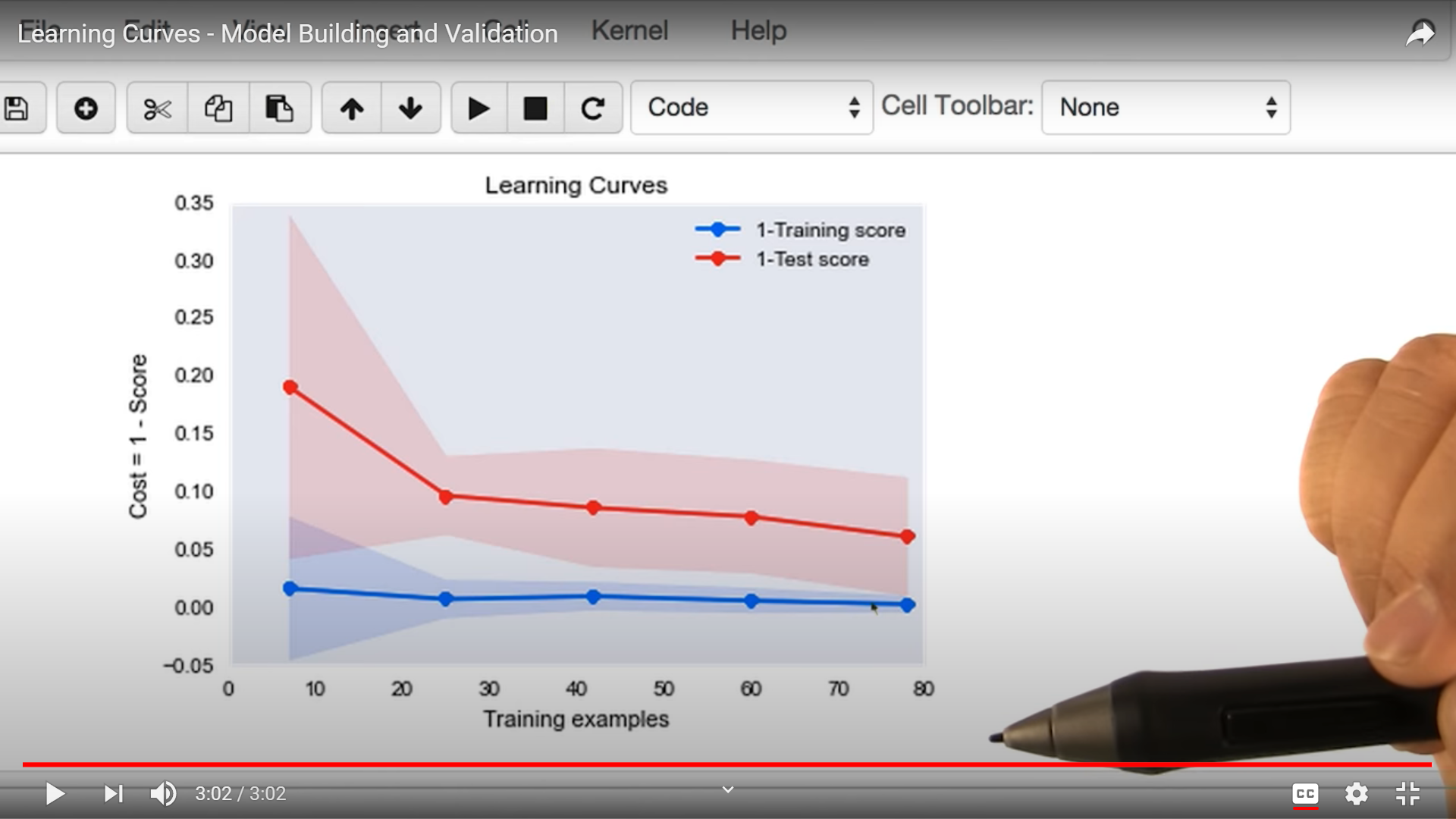


Validation curve:

In order to validate a model we need a scoring function(metrics and scores).

( Cost or score vs size of training and test set.)

Classifiers:score or 1-score.



If you see in this example we have target function as cost and input as training example.after some point of data we see there is no improvement or change in data.so for identifying this learning curve helps us.

In learning curve what we do is we will plot the values of validation or training score between the training samples.we will compare both validation and training score and see how well they are converging.

Actually this learning curve shows us from which data points that the model is looking idle without any increase or decrease on datapoints.