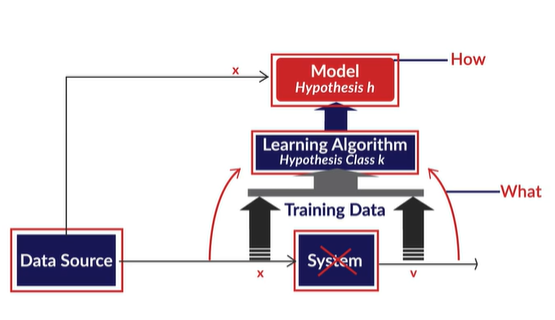
**Model Selection:**



4 unique points about using a simpler model where ever possible:

1. A simpler model is usually more generic than a complex model. This becomes important because generic models are bound to perform better on unseen datasets.
2. A simpler model requires less training data points. This becomes extremely important because in many cases one has to work with limited data points.
3. A simple model is more robust and does not change significantly if the training data points undergo small changes.
4. A simple model may make more errors in the training phase, but it is bound to outperform complex models when it sees new data. This happens because of overfitting.

Complex models have high variance and simple ones have high bias

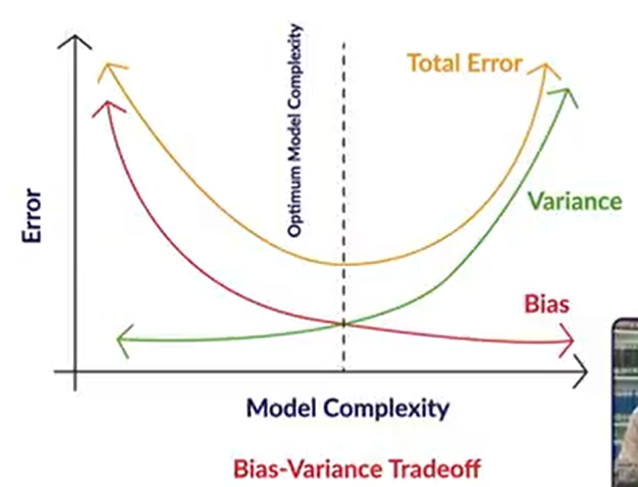
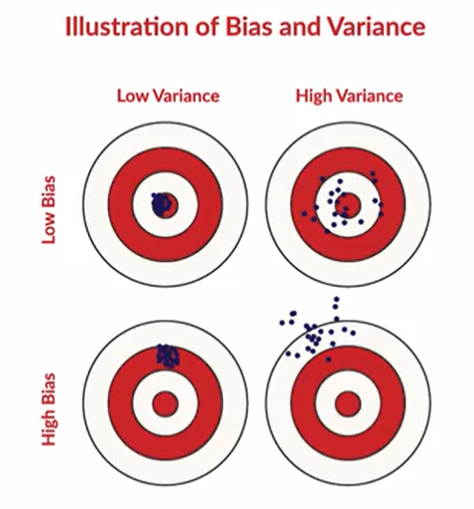
Simple models have high bias and low variance

As the model complexity goes up, the bias reduces while the variance increases, hence the trade-off.

 model (1) has 3 features and model (2) has 11 features. Which model is likely to undergo a larger change when a new training dataset is used? Model 2 - *Model 2 can change its 11 coefficients (and the constant term) to fit the new training data.*

The bias is high when the model is too simple.

Regularization is the process of deliberately simplifying models to achieve the correct balance between keeping the model simple and yet not too naive. Recall that there are a few objective ways of measuring simplicity - choice of simpler functions, lesser number of model parameters, using lower degree polynomials, etc.



**Occam's Razor**

* A model should be as simple as necessary, but no simpler
* When in doubt, choose a simpler model
* Advantages of simplicity are generalizability, robustness, making few assumptions and less data required for learning

**Bias-Variance Tradeoff**

* Bias measures how accurately a model can describe the actual task at hand
* Variance measures how flexible the model is with respect to changes in the training data
* As complexity increases, bias reduces, and variance increases, and we aim to find the optimal point where the total model error is the least

**Overfitting**

* A model memorizes the data rather than intelligently learning the underlying trends in it
* It arises because it is possible to memorize data, and this is a problem, because the real test happens on unseen, real world data

**Regularization** discourages the model from becoming too complex even if the model explains the (training) observations better.

**Hyperparameters** are parameters that we pass on to the learning algorithm to control the complexity of the final model. Hyper parameter are choices that the algorithm designer makes to ‘tune’ the behavior of the learning algorithm.

The various types of cross-validation are -

* K-fold cross-validation
* Leave one out(LOO)
* Leave P-out(LPO)
* Stratified K-Fold

<https://scikit-learn.org/stable/modules/cross_validation.html>