Model Evaluation

Human Learning Evaluation

- Teacher explains logic with some problem, how can we test whether a student grab the logic or not?

- Goal: The logic should be tested such a way that (s)he could solve any new variation of that problem in future.

- Which aspects of learning we wish to evaluate?

  Bias & Variance of learning

Human Learning Evaluation

- What aspects of learning teacher wishes to test?
  - Bias: How accurately (s)he can solve different variations of problem. In other words, the extent to which his/her averaged accuracy differs from expected accuracy
  - Variance: How much consistent in his/her capability of solving those variations of problem

Human Learning Evaluation

- Approach1 (Resubstitution): Give him/her same problem and check!!

- Approach2 (Resampling): Give different variations of problem and check!! There are multiple ways of doing this.
  - Generate variations of problem by replacing just numbers, by rewrite of problem, by making objective version, etc.,
  - Each approach measures bias & variance of learning slightly different way

Machine Learning(Model) Evaluation

- How do you evaluated machine learned model?
- Goal: Select the model that best performs on unseen(future) data.
- Which aspects of learning we wish to evaluate?

Bias & Variance of model

Approach1: Resubstitution Error

- Use entire train data for learning as well as evaluation
- The error returned by model on same training set used for learning is called as Resubstitution error.
- Does this approach makes sense???

Issues with Resubstitution Error

- Model may not have enough data to fully learn the concept (but on training data we don't know this)

- For noisy data, the model may overfit the training data

- Hence, Resub error is always too optimistic i.e., under-estimates true error of model

Approach2: Resampling Error

- Hold some observations from train-set, called as validation set

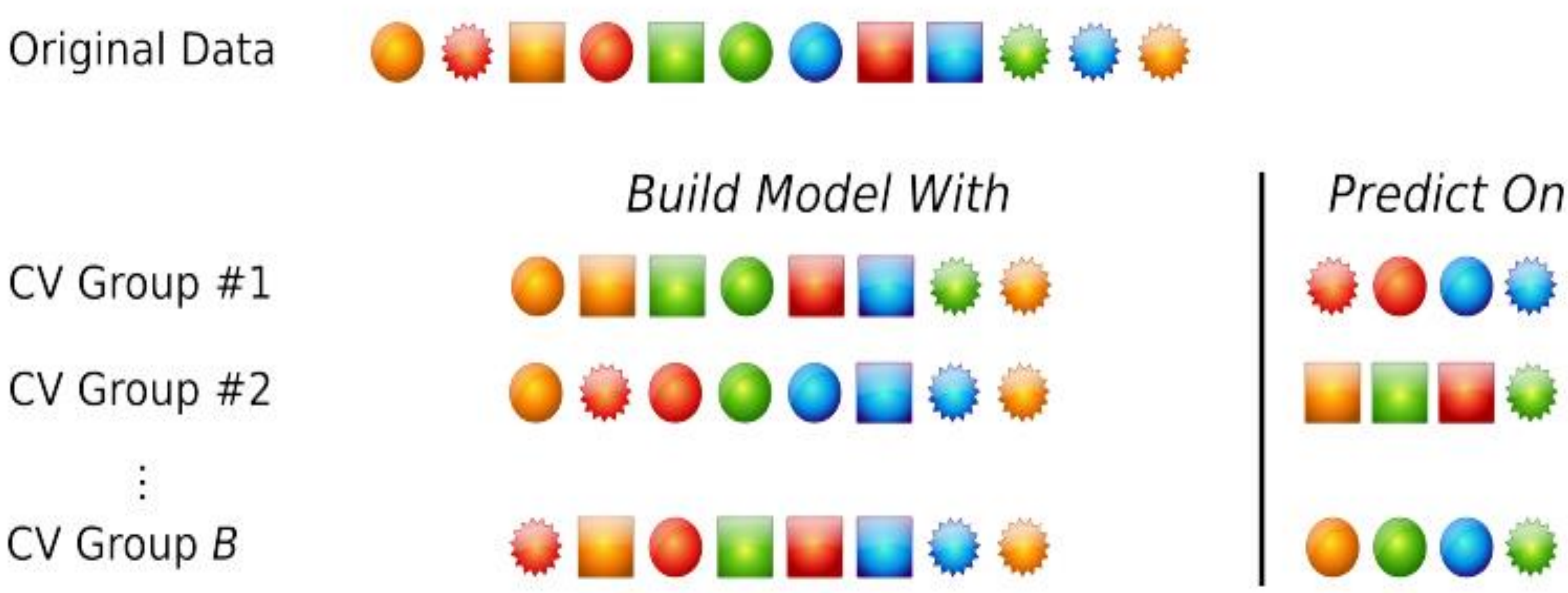
- Use a validation set to estimate how well the model perform on new unseen data(out of sample)

- Resampling methods try to “inject variation” in the system to approximate the model’s performance on future samples.

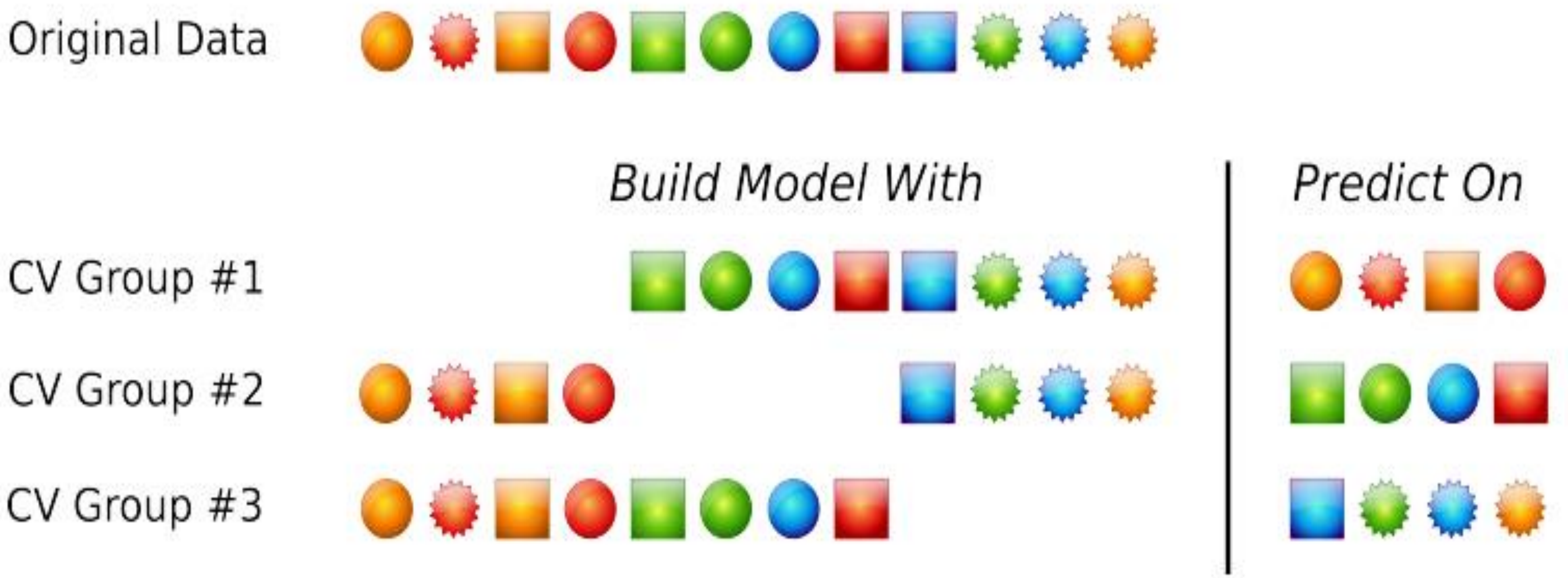
Resampling Methods

- Repeated Holdout(with stratification)
- Cross Validation(with stratification)
- Bootstrapping

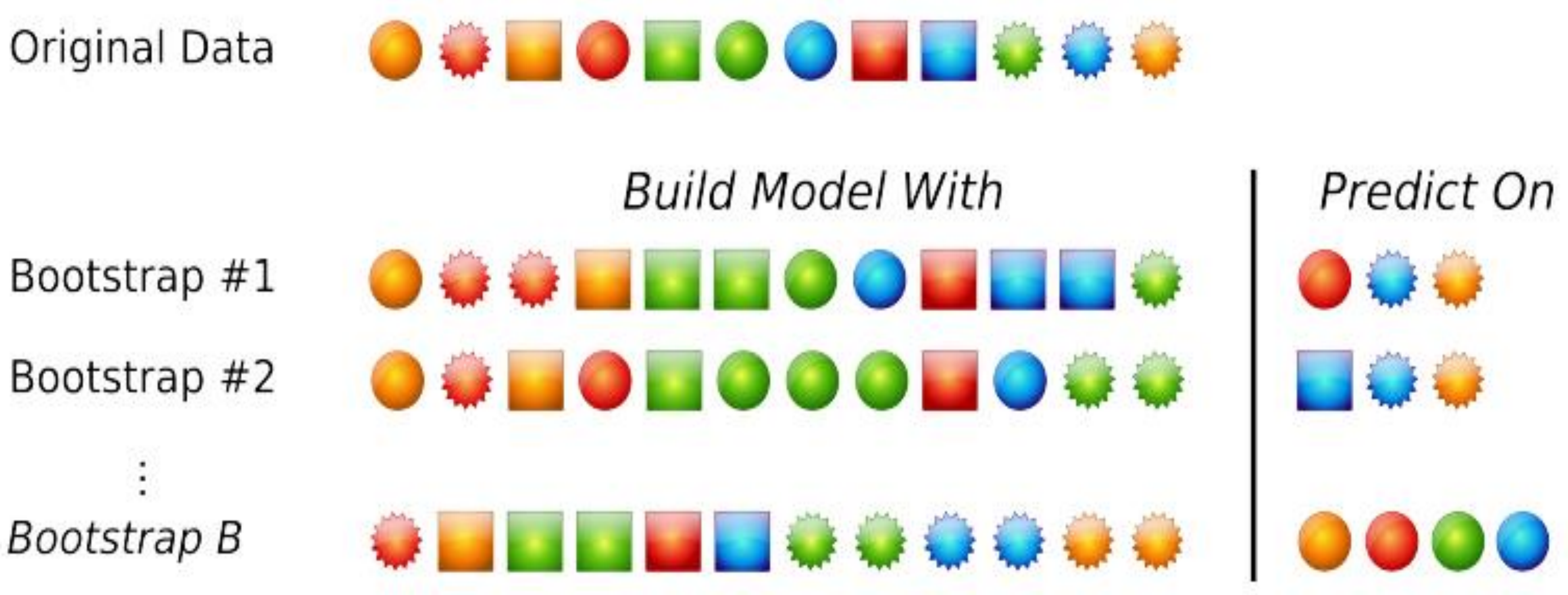
Repeated Holdout Illustrated



K-Fold CV Illustrated



Bootstrapping Illustrated



Parameter Tuning

Parameter Tuning Problem

- Which parameter values do you select for given model?

- Goal: Select the parameters of model that provides best performance on unseen(future) data.

Parameter Tuning Algorithm

- Define sets of model parameter values to evaluate

- For each parameter set do
  - Estimate the average error/accuracy of model using chosen resampling technique

- Choose the parameters that gives smallest error or highest accuracy, and build the model on entire data with the chosen parameters.

Model Selection

Model Selection Problem

- If you have multiple models, which model do you select for deployment?

- Goal: Select the model that best performs on unseen(future) data.

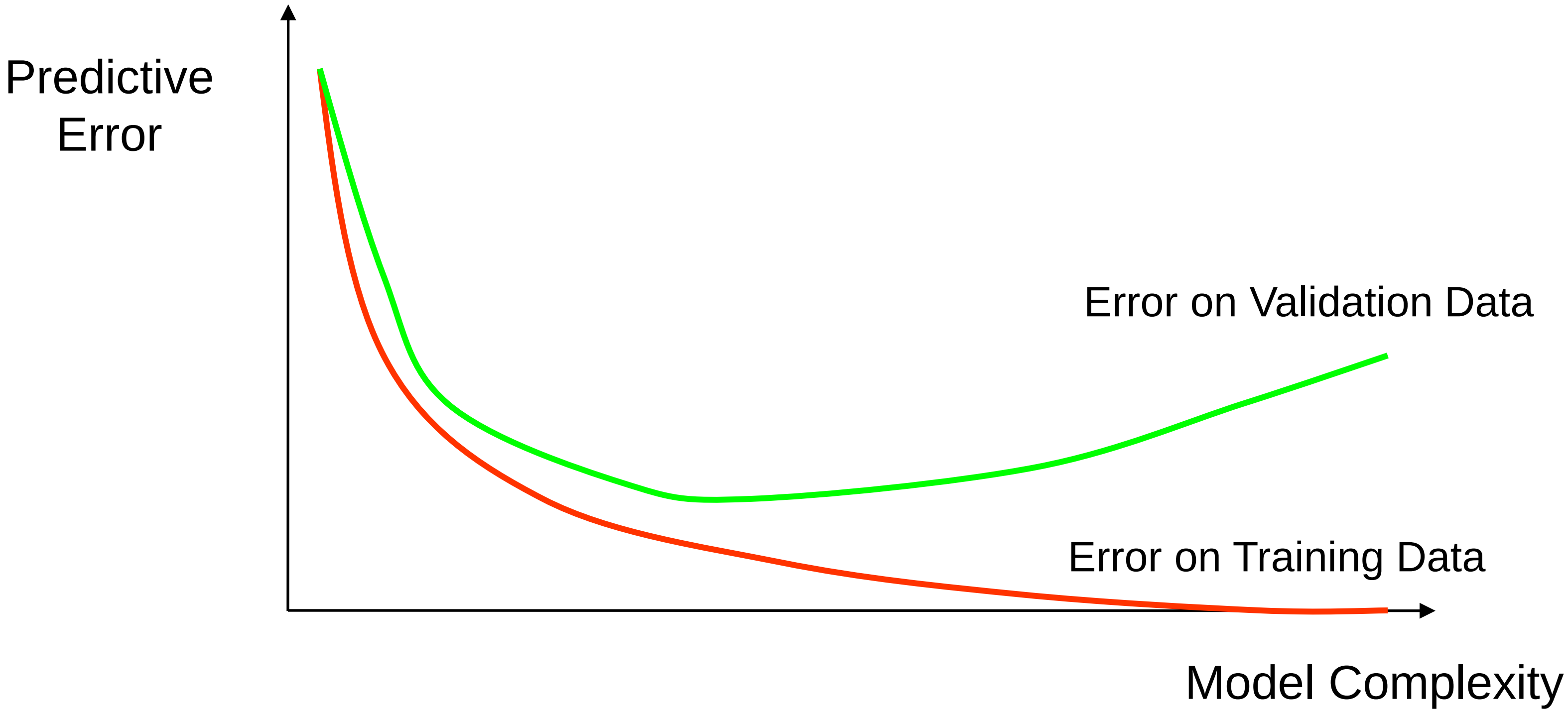
Model Selection Algorithm

- Define the bunch of models we wish to evaluate

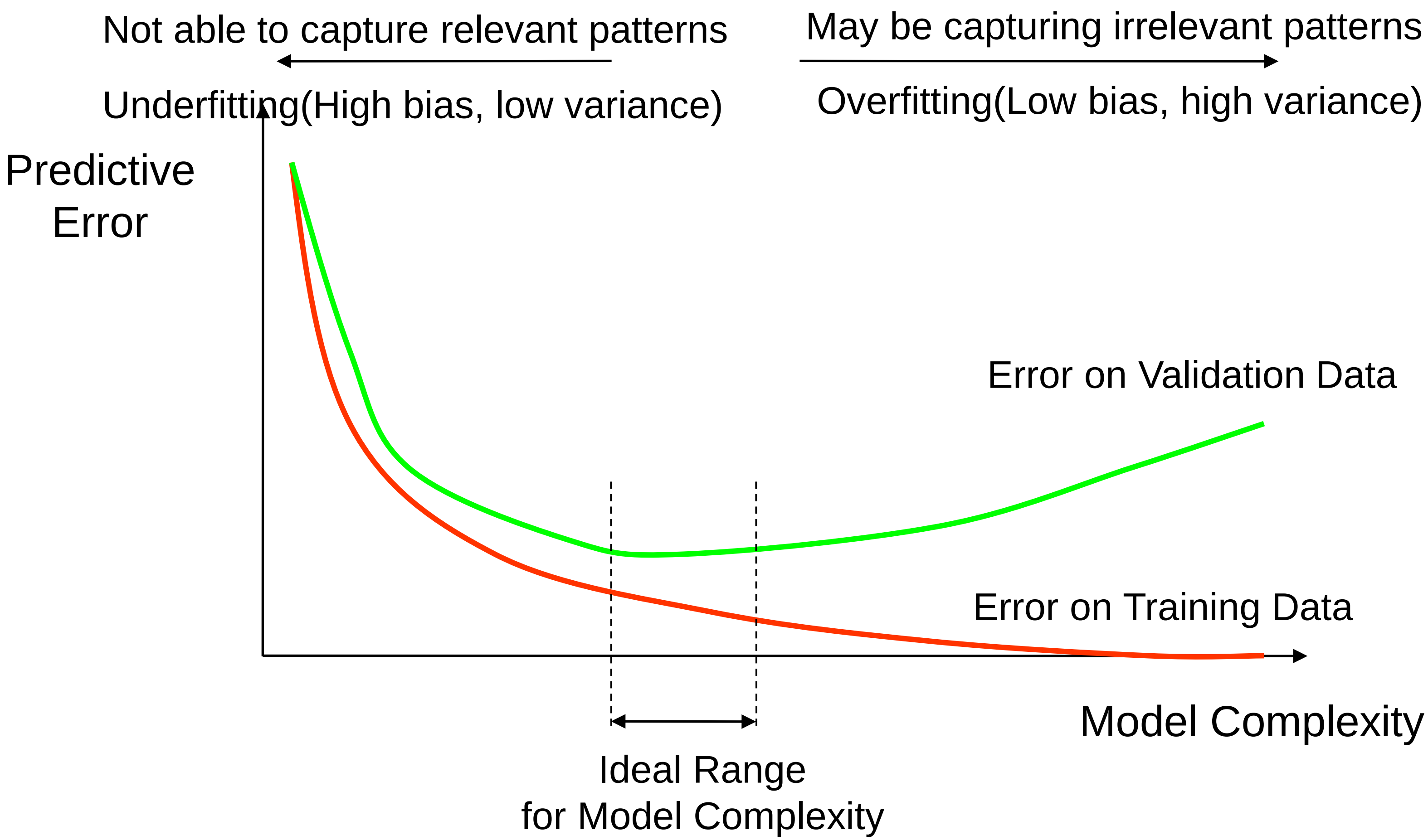
- For each model do
  - Estimate the average error/accuracy of model using chosen resampling technique

- Choose the model that gives smallest error or highest accuracy.

Model Selection Problem



Model Selection Problem



Simpler Models are not able to capture relevant patterns and might have too big of a bias in predictions.

Complex Models may "chase" irrelevant patterns in the training data that are not likely to exist in future data.