Model Selection & Parameter Tuning

Model Selection Problem

- Which model do you select for future(test data) predictions?

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

Parameter Tuning Problem

- Which parameters of model do you select for future(test data) predictions?

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

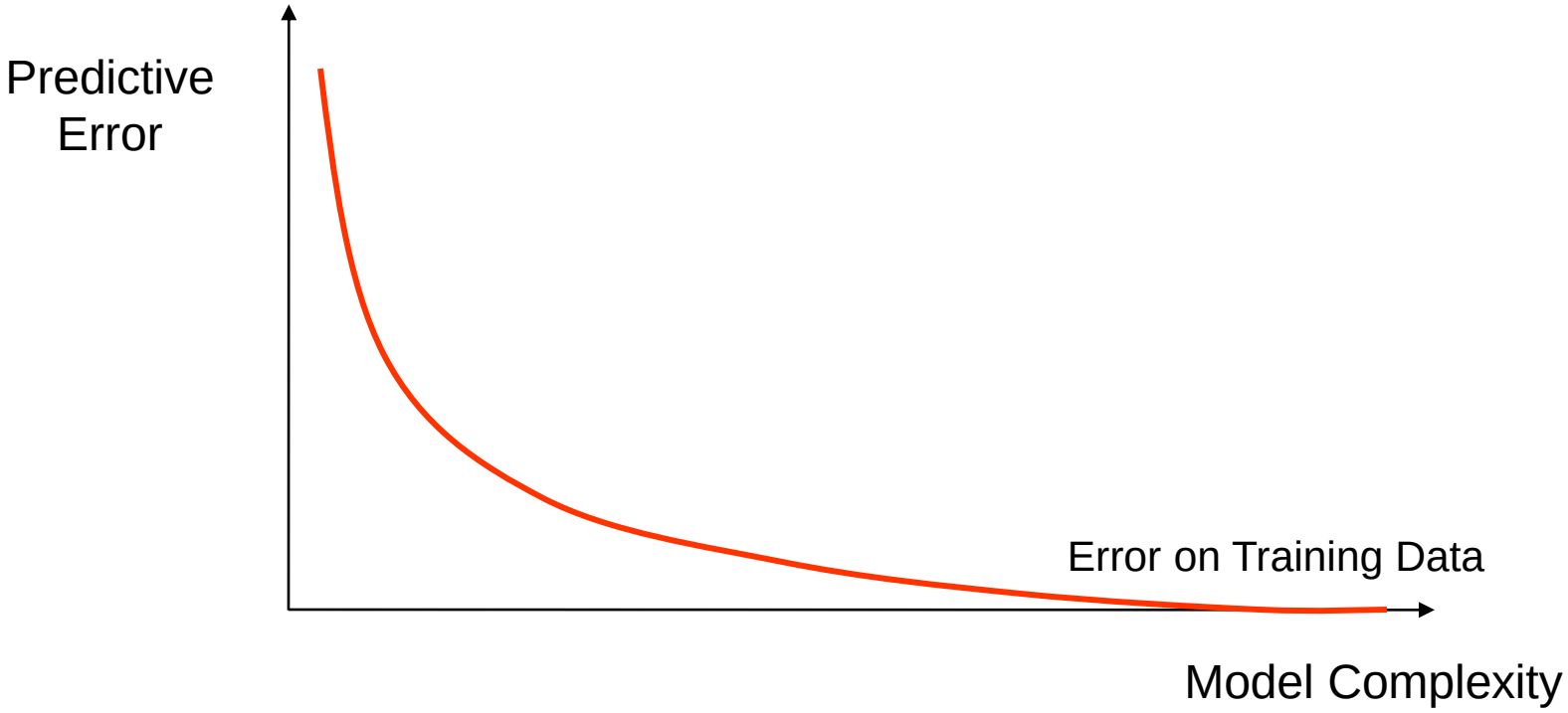
Solution1: Resubstitution Error

- Use entire train data for learning as well as evaluation
- Select the model or parameter values that gives smallest Resubstitution error
- Does this approach makes sense???

Solution1: Resubstitution Error

- Issues
  - 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

Model Complexity vs Training Error



Solution2: Resampling Error

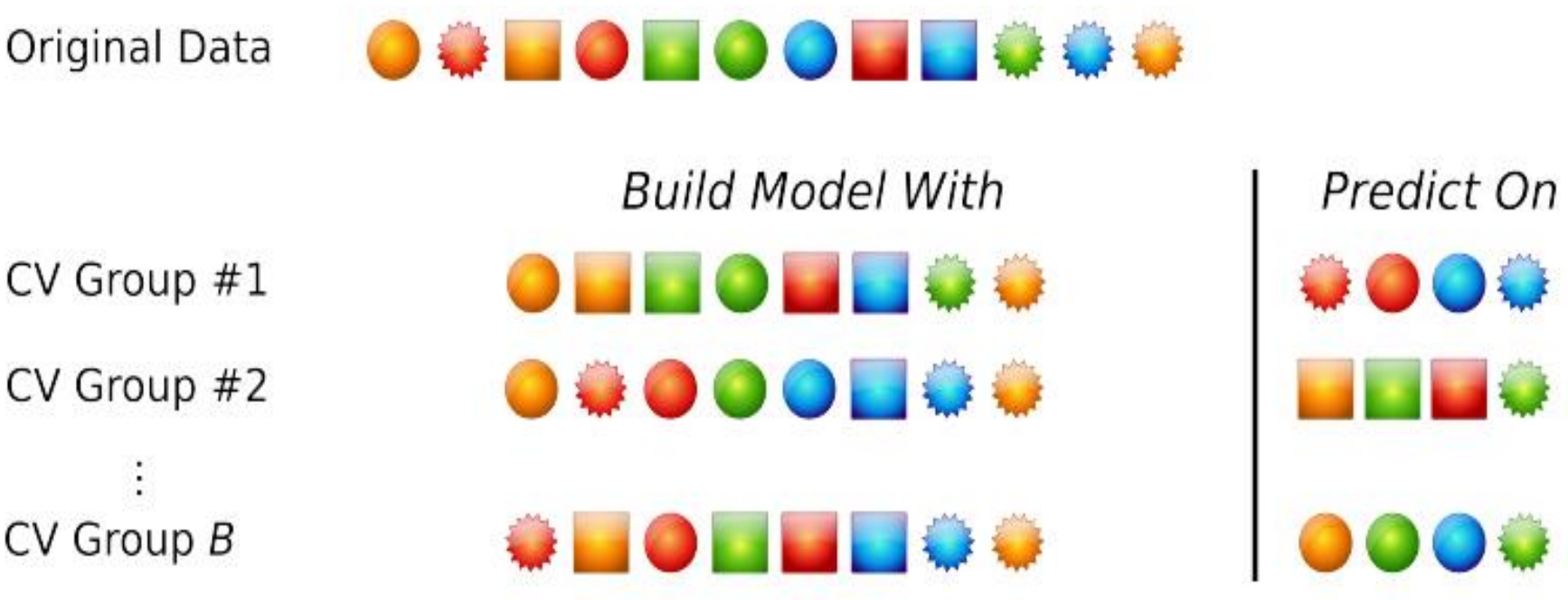
- Use a validation set to estimate how well the model perform on new unseen data(out of sample)?
    - This is the true test of what we have learned (just like a classroom)

- Resampling methods try to "inject variation" in the system to approximate the model's performance on future samples.

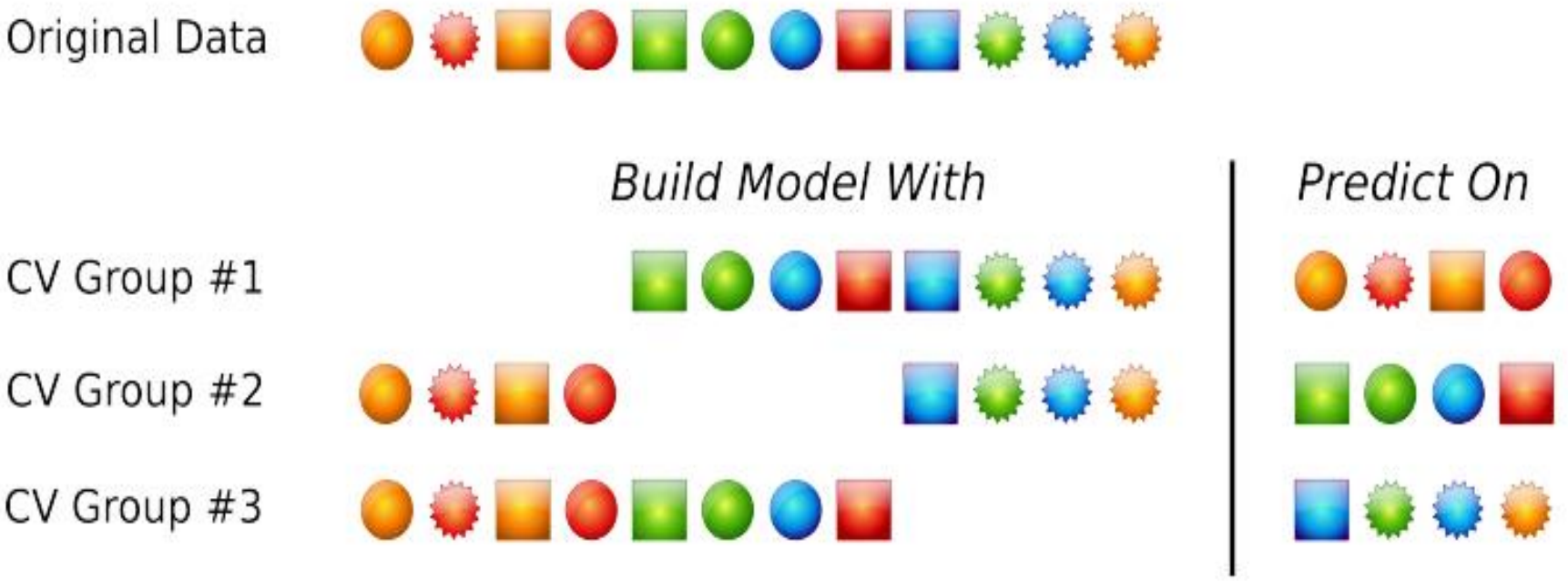
Resampling Methods

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

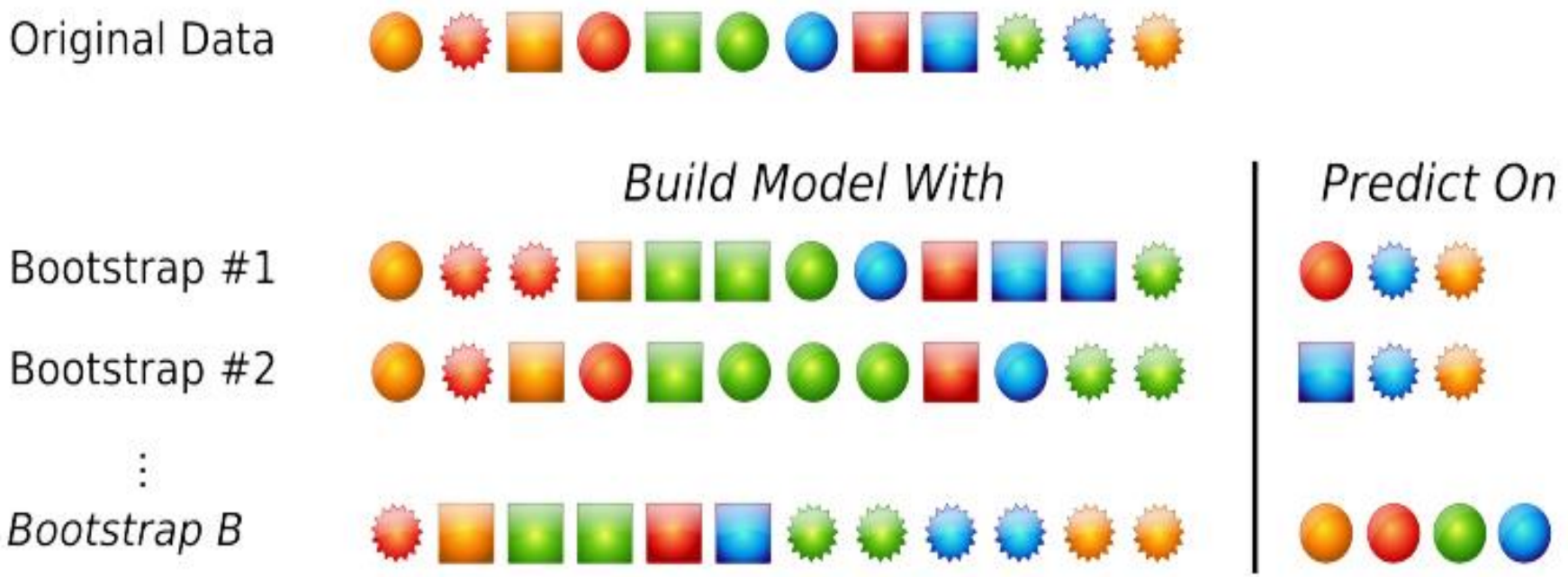
Repeated HoldOut



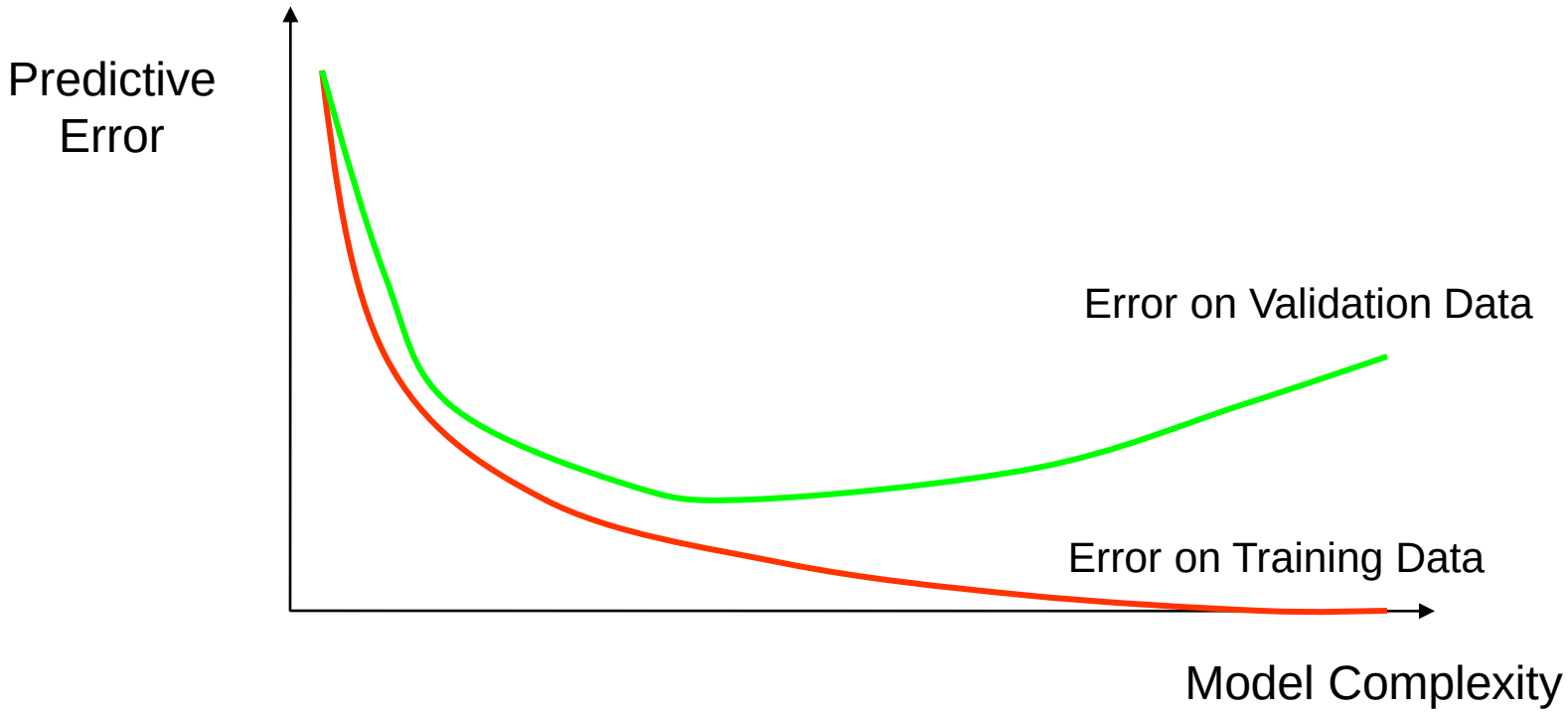
K-Fold Cross Validation



Bootstrapping



Model Complexity vs Validation Error



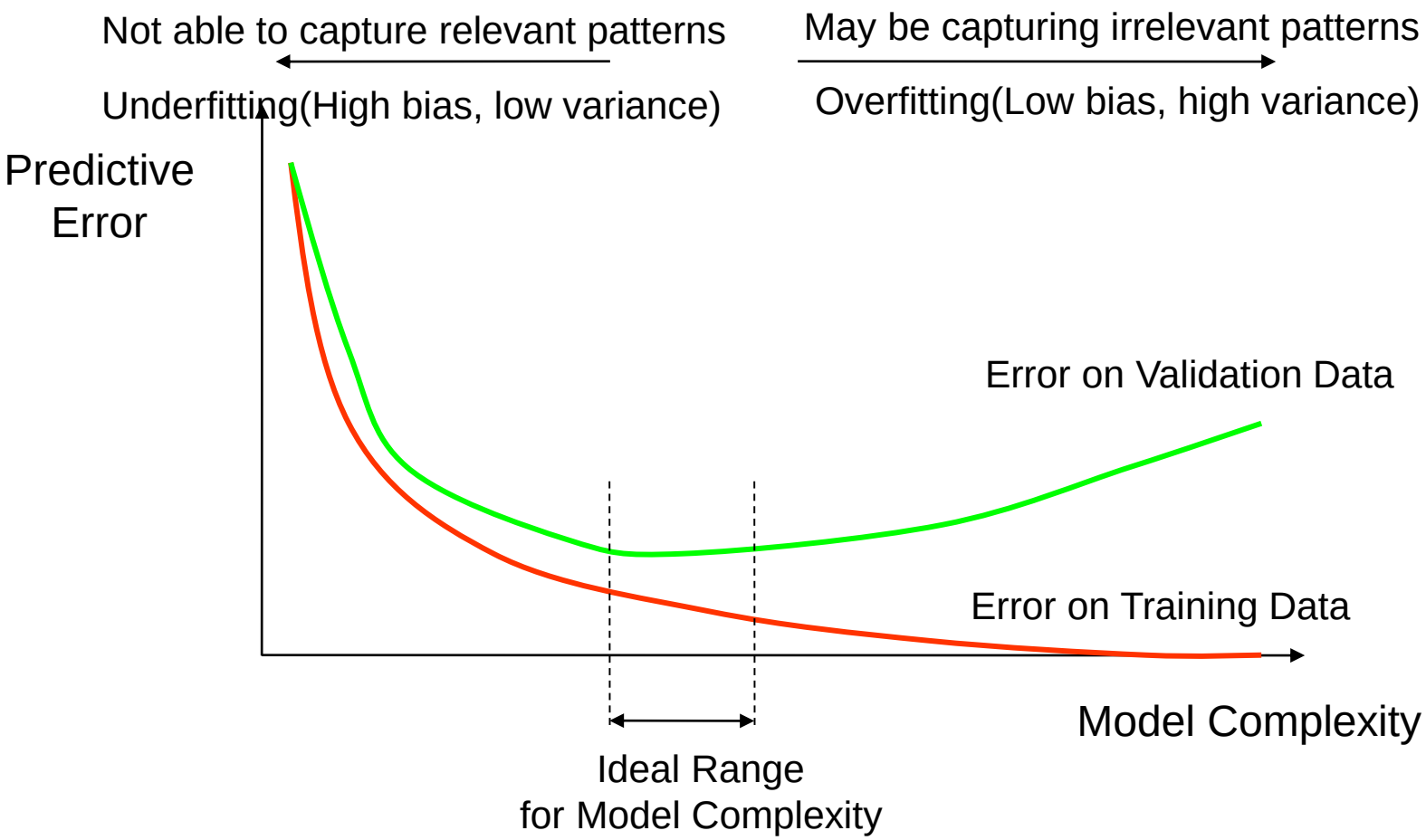
Model Selection Criteria

Goal : Models that captures important patterns of the data that generalize to the future observations we are trying to predict.

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.

Model Selection



Model Selection

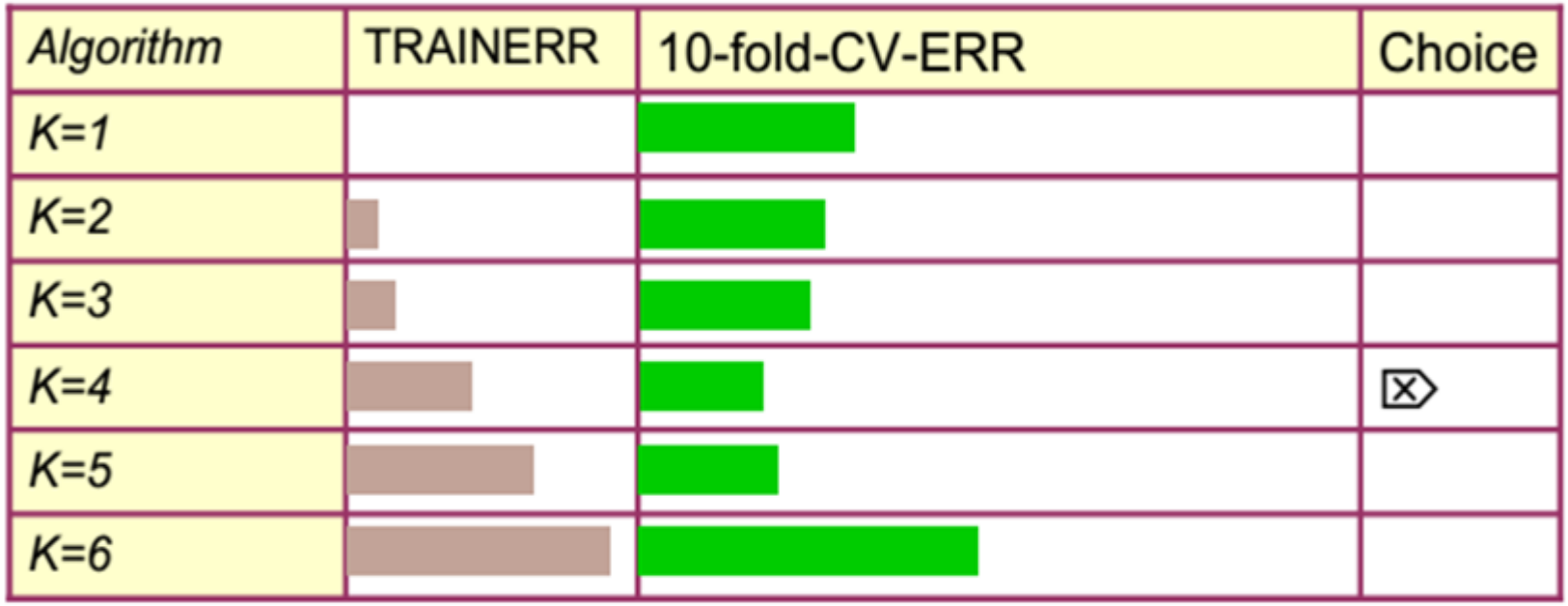
- Example: Choosing which regression algorithm to use
- Step 1: Compute 10-fold-CV error for six different model classes:

Algorithm	TRAINERR	10-fold-CV-ERR	Choice
1-NN			
10-NN			
Linear Reg'n			
Quad reg'n			⊠
LWR, KW=0.1			
LWR, KW=0.5			

- Step 2: Whichever algorithm gave best CV score: train it with all the data, and that's the predictive model you'll use.

Parameter Tuning

- Example: Choosing "**cp**" for a **Decision Tree Learning**
- Step 1: Compute CV error for **different values of cp**



Algorithm	TRAINERR	10-fold-CV-ERR	Choice
K=1			
K=2			
K=3			
K=4			⊠
K=5			
K=6			

- Step 2: Whichever model class gave best CV score: train it with all the data, and that's the predictive model you'll use.

Parameter Tuning

```
Define sets of model parameter values to evaluate;
for each parameter set do
    for each resampling iteration do
        Hold–out specific samples ;
        Fit the model on the remainder;
        Predict the hold–out samples;
    end
    Calculate the average performance across hold–out predictions
end
Determine the optimal parameter set;
```