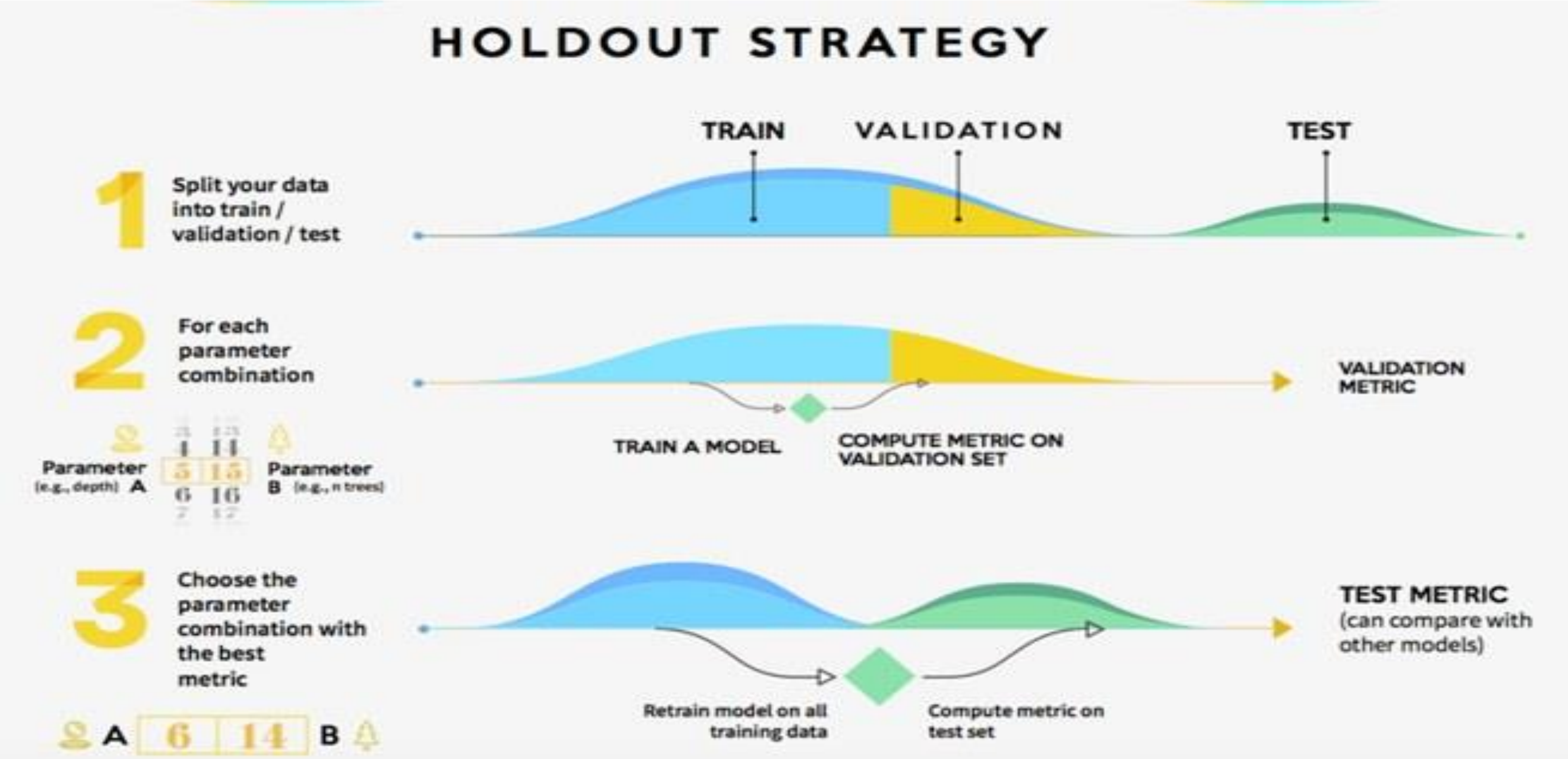
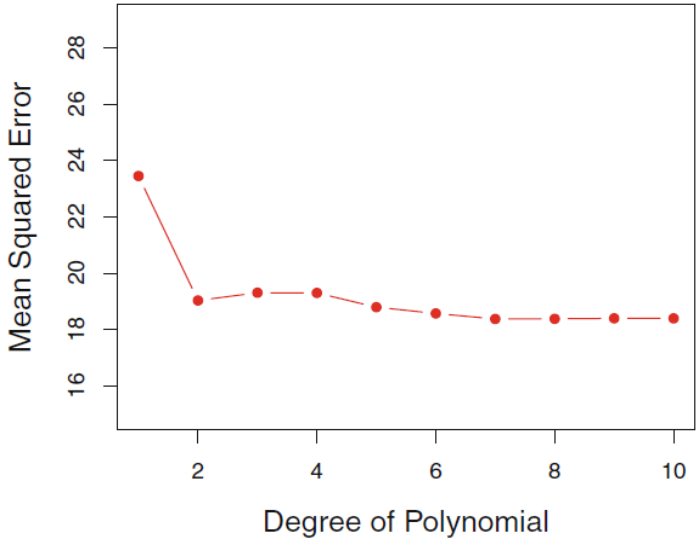
Evaluation of Classifiers

By
Vinod P. (Ph.D, Computer Engg.)

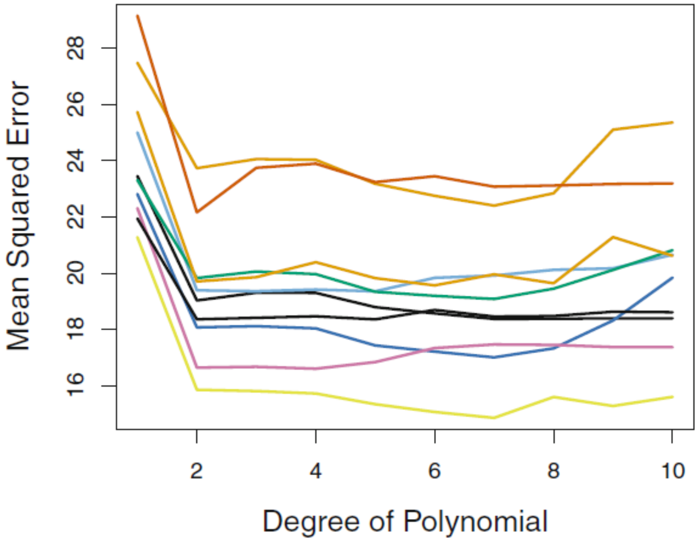
Holdout Cross-validation



Holdout

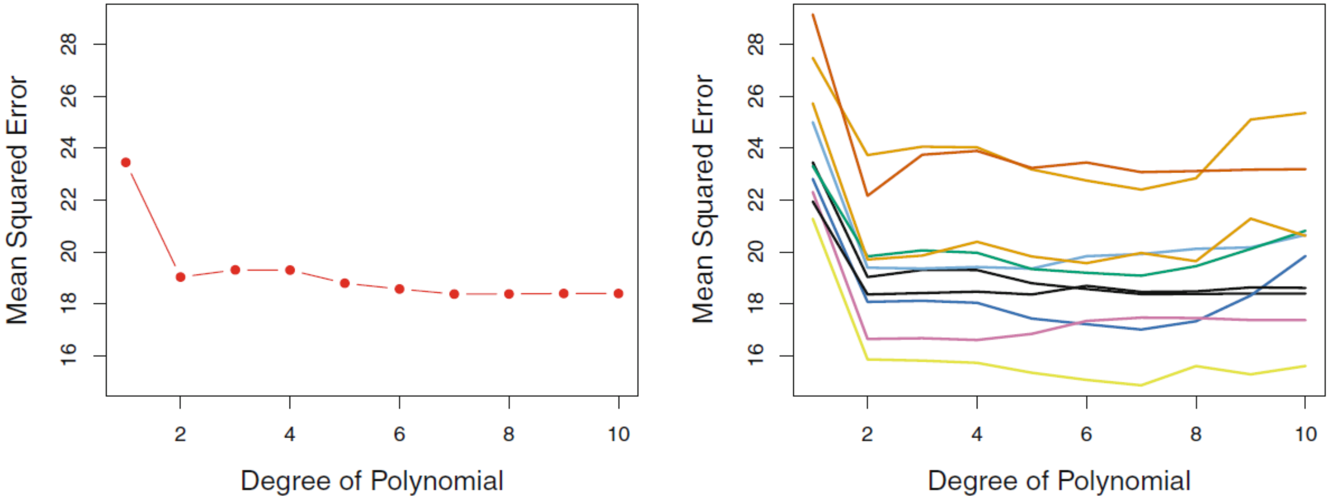


Randomly divide data into training set and Validation set



Ten different randomly split of training set and Validation set

Holdout



No consensus among the curves as to which model results in the smallest validation set MSE !!!!

Drawbacks-Validation set approach

- Test error can be highly variable depending on observations included in the training set and validation set

- Subset of observations i.e., training set are used to fit the statistical model
  - Since the model performs worse when trained on fewer observations, suggest that validation set error overestimates the test error rate

Holdout Cross-validation

- The limitations of the holdout can be overcome with a family of resampling methods at the expense of more computations
  - Cross Validation
    - Random Subsampling
    - K-Fold Cross-Validation
    - Leave-one-out Cross-Validation
  - Bootstrap

What is Resampling?

- Repeatedly draw samples from training set and refit the model of interest to obtain additional information about the fitted model
- Example: Estimate the variability of linear regression
    - Draw samples from the training data
    - Fit linear regression to each new sample
    - Examine the extent to which the results differ

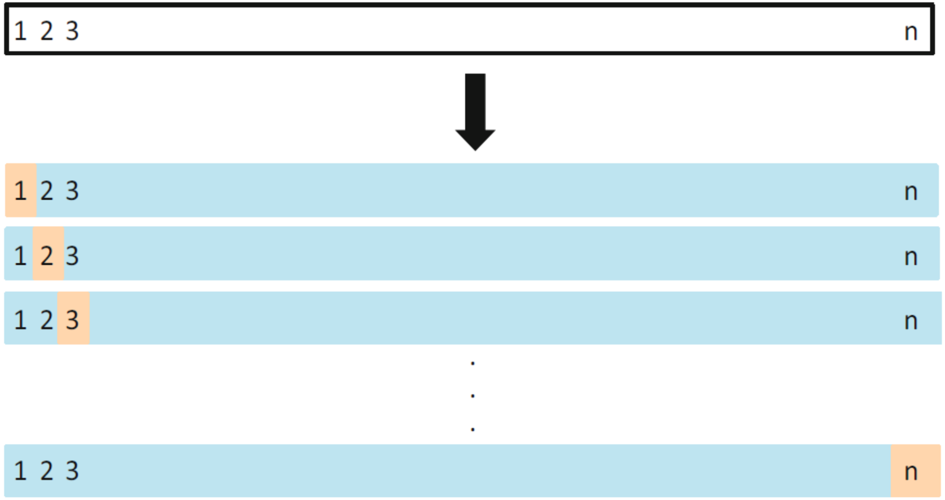
General Terms

- Model assessment: Process of evaluating model's performance
- Model selection: Process of selecting the flexibility for a model

Why cross-validation?

- Test error: Predict the response of a statistical learning model on new observation
- Test error can be calculated if the test set is available
- Training error and test error are often different
- Cross-validation
  - Used in the absence of large test set to estimate test error
  - Set of techniques used to estimate the quality of model using the training set

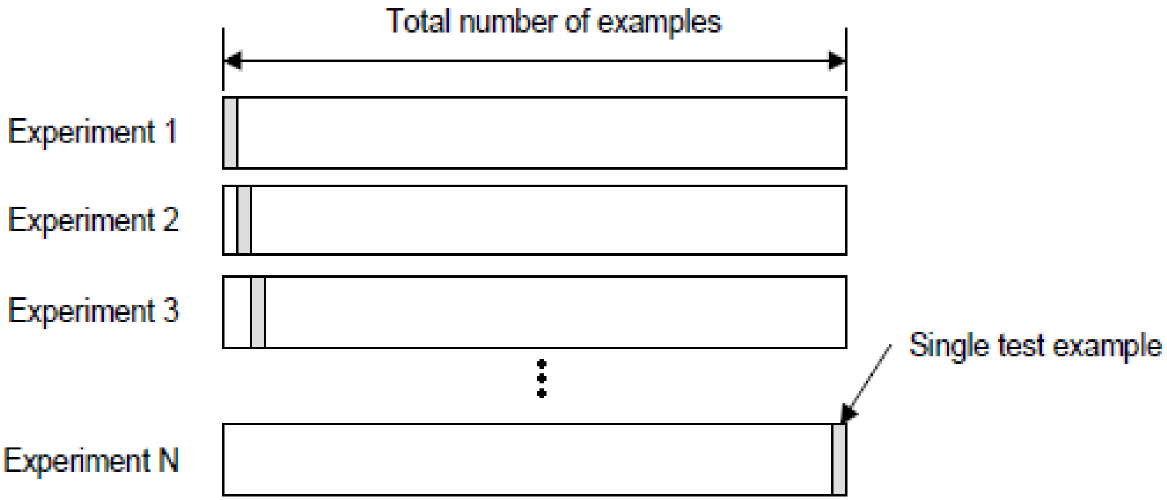
Leave-one-out Cross-validation



- A single observation **(x1, y1)** is used for the validation set
- Statistical model is fit on (*n-1*) training set, prediction $\hat{y}_1$ is made

$$\{(\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2) \ldots, (\mathbf{x}_n, \mathbf{y}_n)\}$$

Leave-one-out Cross-validation

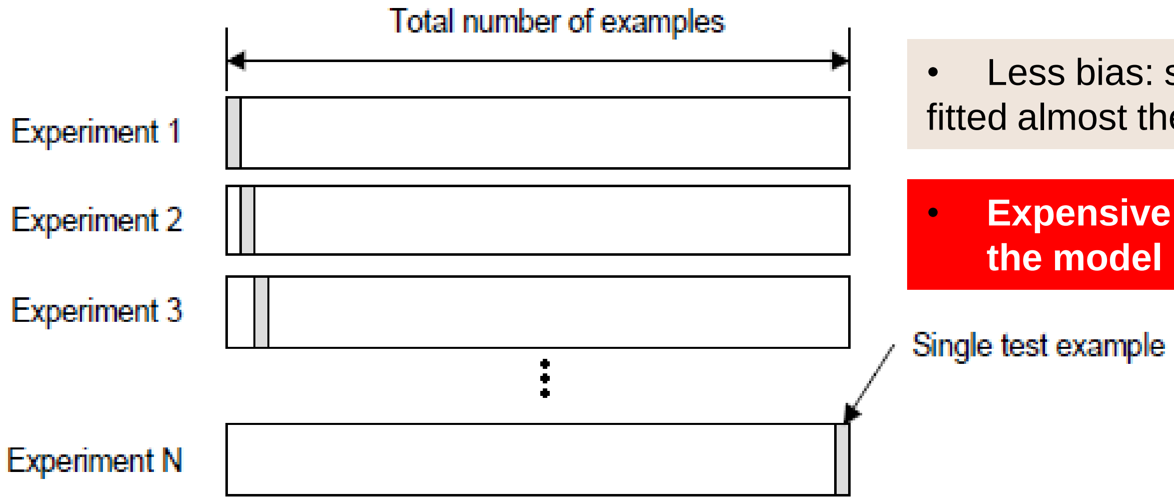


- True error is estimated as the average error rate on test examples

$$E = \frac{1}{N}\sum_{i=1}^{N} E_i$$

Error $E_i = (y_i - \hat{y}_i)^2$

Leave-one-out Cross-validation



- Less bias: statistical model is fitted almost the entire data

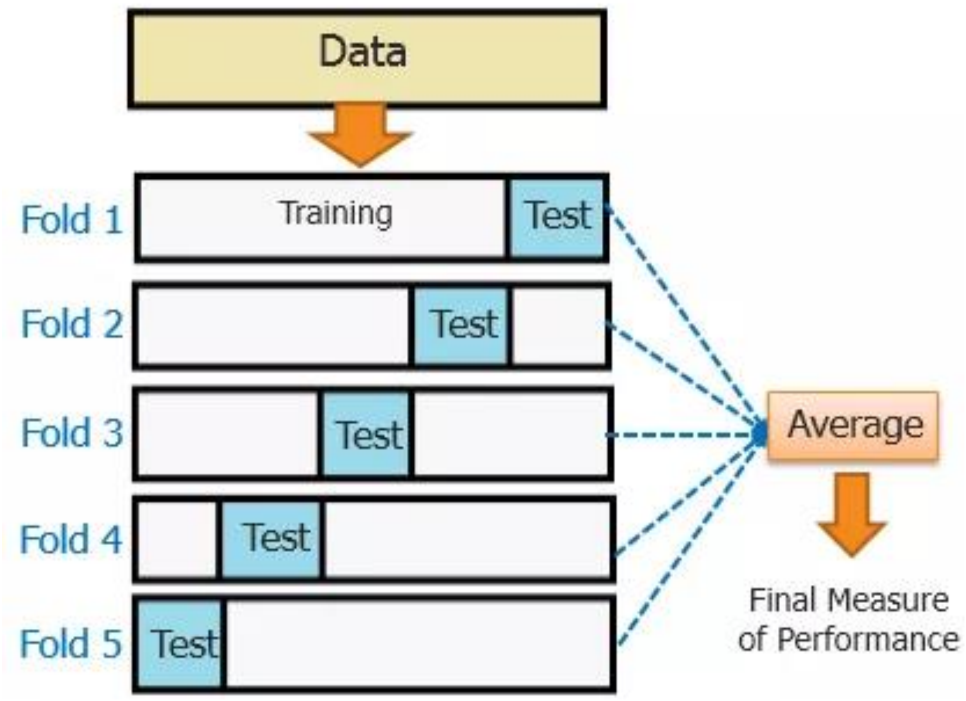
- **Expensive to implement since the model has to fit *n* times**

- True error is estimated as the average error rate on test examples

$$E = \frac{1}{N}\sum_{i=1}^{N} E_i$$

Error $E_i = (y_i - \hat{y}_i)^2$

Cross-validation



Cross-validation is a technique to evaluate the predictive models by partitioning the original sample into a training set to train the model, and a test set to evaluate it.

Simplest method: Holdout

K-fold Cross-validation

- Randomly divide the set of observations into *k* groups or *folds*, approximately of equal size
- First fold is treated as validation set, fits the remaining *(k-1)* folds
- Mean square error $\mathbf{MSE}_1$ is computed for observations in held-out fold

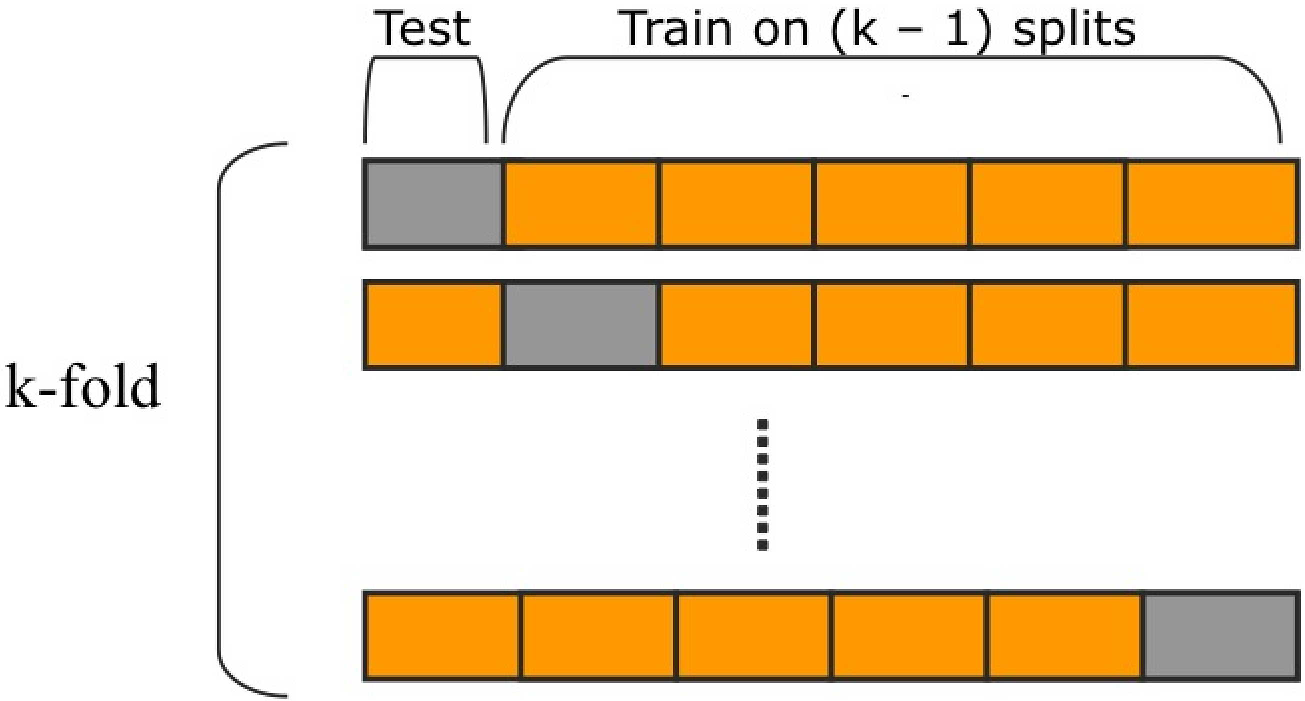
K-fold Cross-validation

- Procedure is repeated $\boldsymbol{k}$ times, each time a different group of observation is treated as validation set

- The k-fold CV estimate is computed by averaging MSE’s

$$CV_{(k)} = \frac{1}{k}\sum_{i=1}^{k} MSE_i$$

K-fold Cross-validation

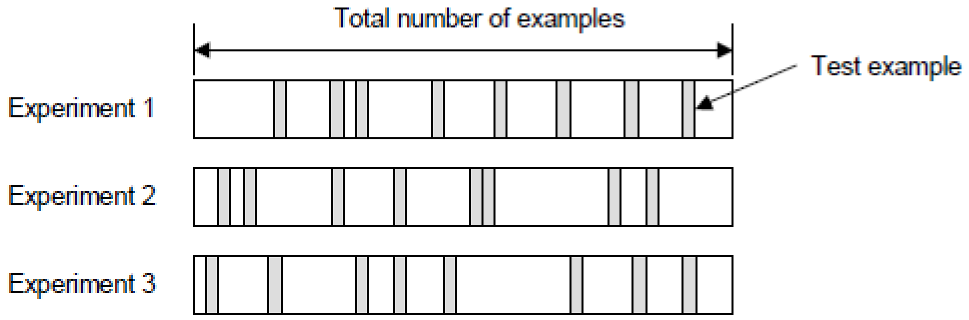


10-fold Cross-validation

$$E = \frac{1}{10}\sum_{i=1}^{10} E_i$$

Random Sampling (1/2)

- Performs *K* splits of the dataset
  - Each split randomly selects a (fixed) no. examples without replacement
  - For each data split we retrain the classifier from scratch with the training examples and estimate E_i with the test examples

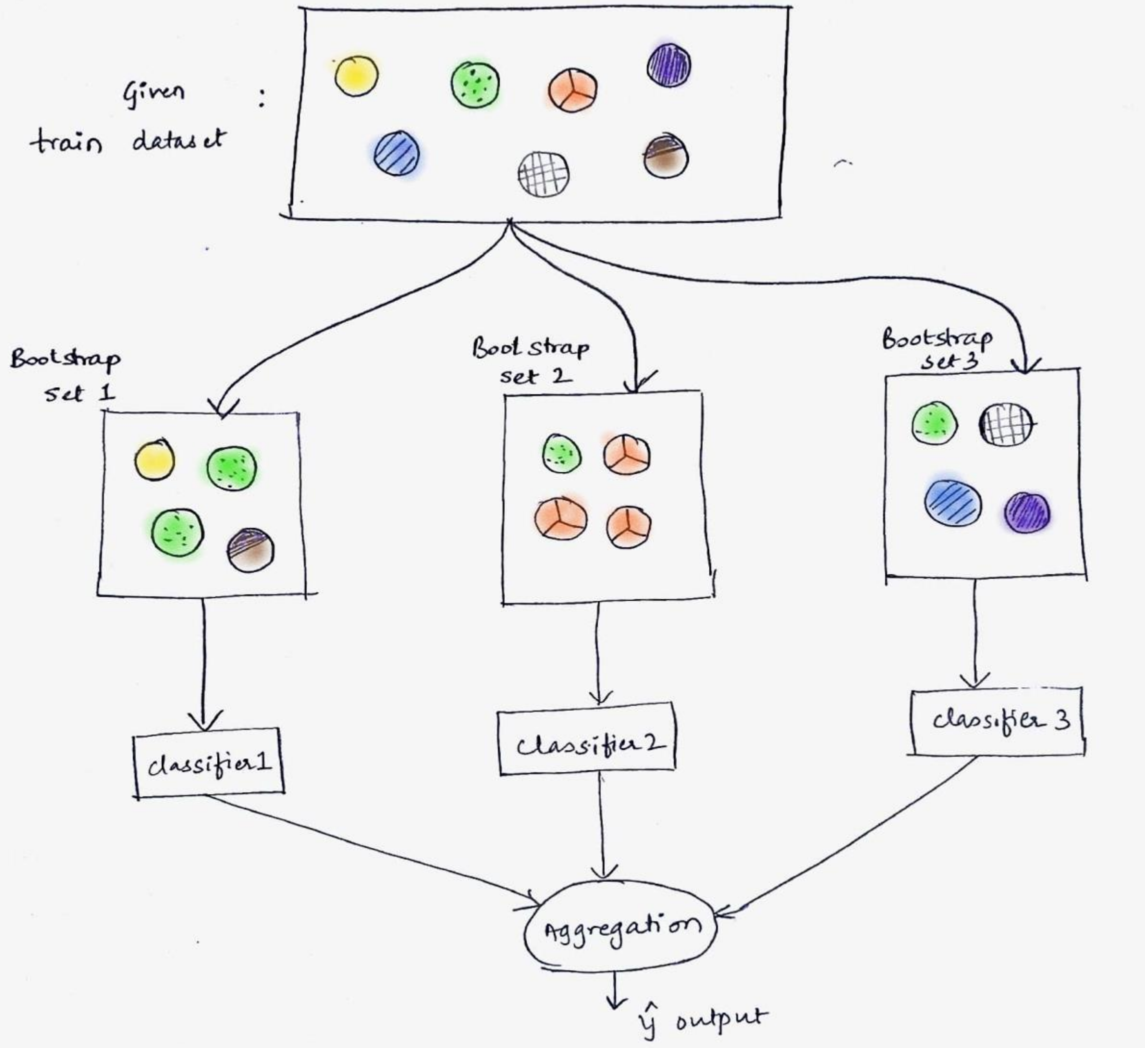


Random Sampling (2/2)

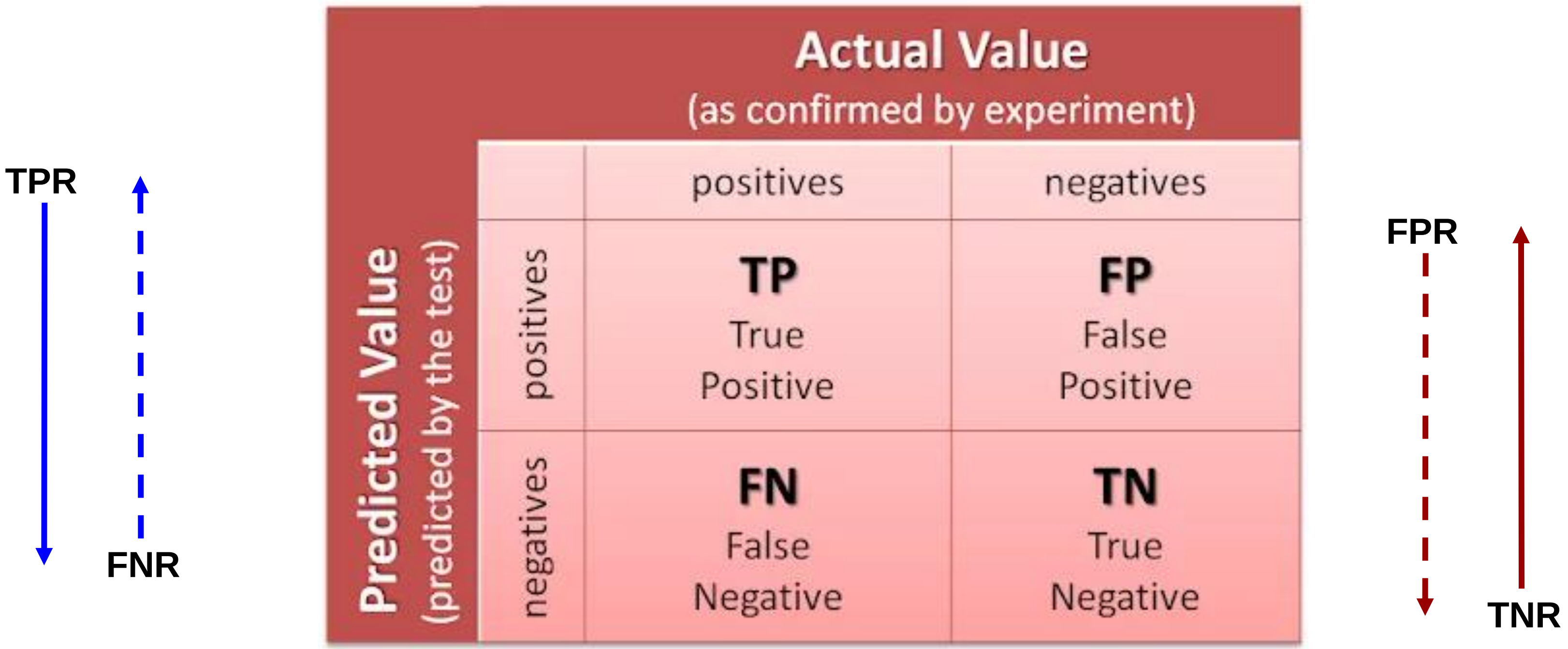
- The true error estimate is obtained as the average of the separate estimates E_i
  - This estimate is significantly better than the holdout estimate

$$E = \frac{1}{K} \sum_{i=1}^{K} E_i$$

Bootstrapping



Confusion Matrix



Confusion Matrix

- True Positive Rate, Sensitivity, Recall or Hit Rate

$$TPR = TP / (TP+FN) = 1-FNR$$

- True Negative Rate, Specificity, or Selectivity

$$TNR = TN / (TN+FP) = 1-FPR$$

- Precision: $P = TP / (TP+FP)$
- Accuracy:

$$(TP+TN)/ (TP+FN+TN+FP)$$

ROC Curve