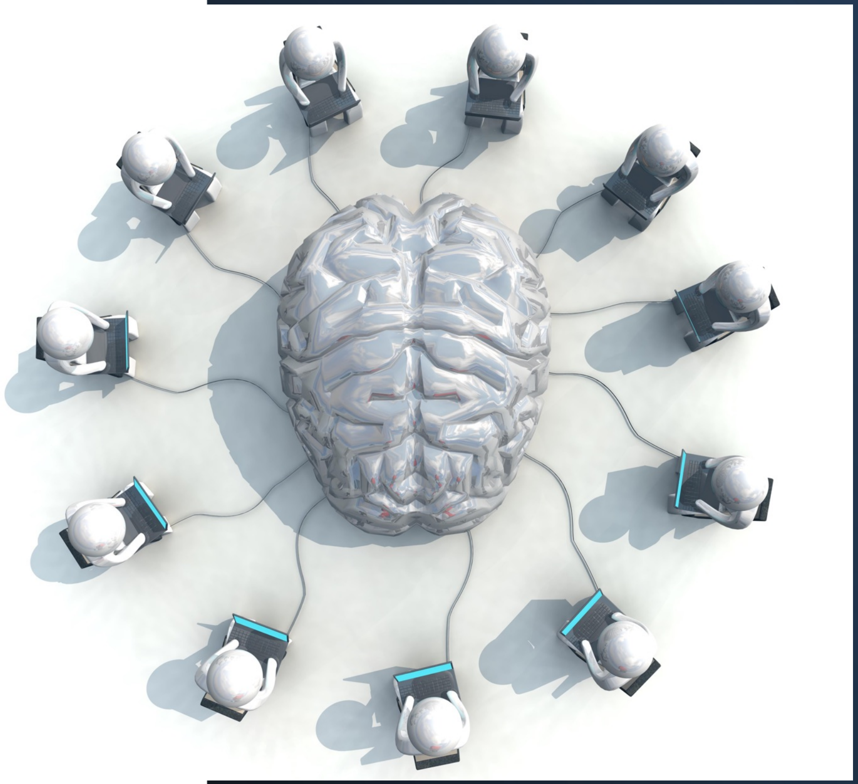
Ensemble Learning

- **Bootstrap Aggregation (Bagging)**
- **RF algorithm**
- **Boosting and others**

Wisdom of Crowds

The collective knowledge of a **diverse and independent** body of people typically exceeds the knowledge of any single individual, and can be harnessed by voting.

- James Surowiecki (2004)



Ensemble learning

- Probable outcomes of developing machine learning models
  - Sometimes weak and inaccurate
  - Some performs better on specific occasions.
- Ensemble learning
  - Generating and combining multiple models (classifiers or experts) to solve a particular computational intelligence problem.

Ensemble learning...

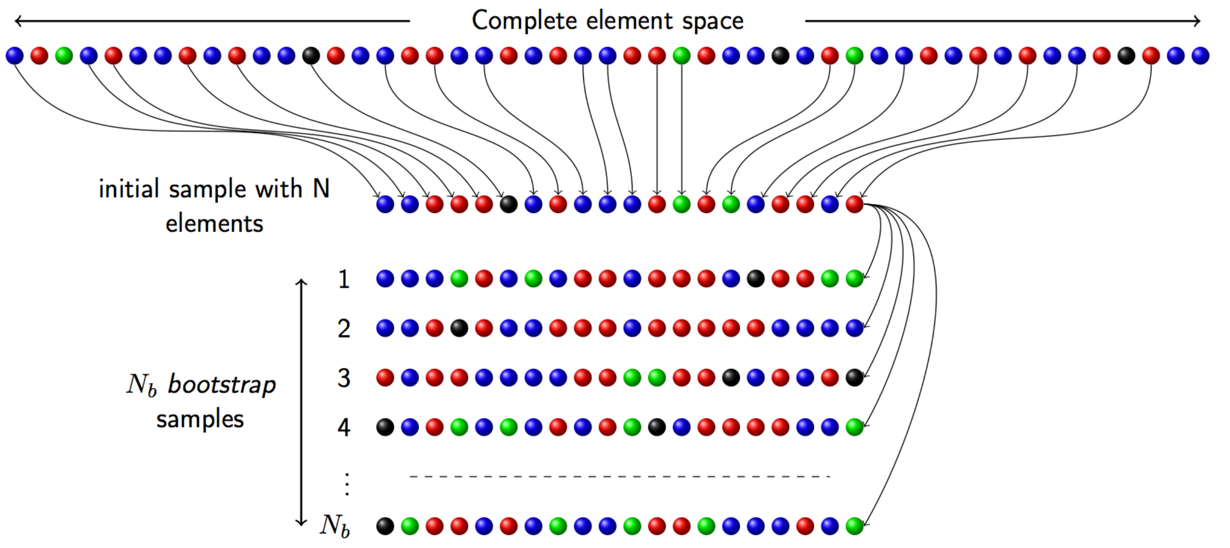
- Consider this scenario:
  - We know that a single decision tree might not perform well
    - It is super-fast to train.
  - What if we create multiple trees?
    - We must make sure that they do not all learn the same thing.

Ensemble learning...

- Problem with single decision tree
  - Risk of overfitting or increased variance
- To reduce the variance of unstable learning methods (such as DT) use ensemble method
  - Train multiple decision trees, each with slightly different subsets of data
  - Take their combined decision
    - Classification
      - Voting
    - Regression
      - Averaging
- Random Forest
  - Ensemble of DT

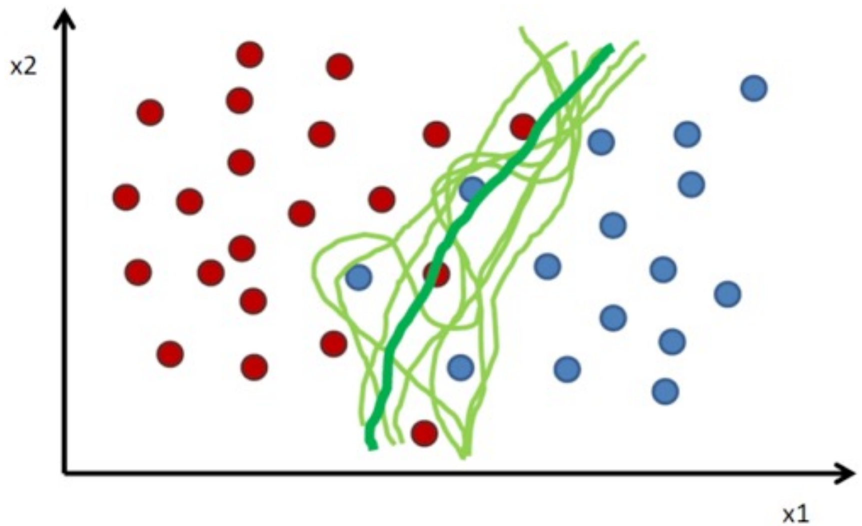
Bootstrap estimation

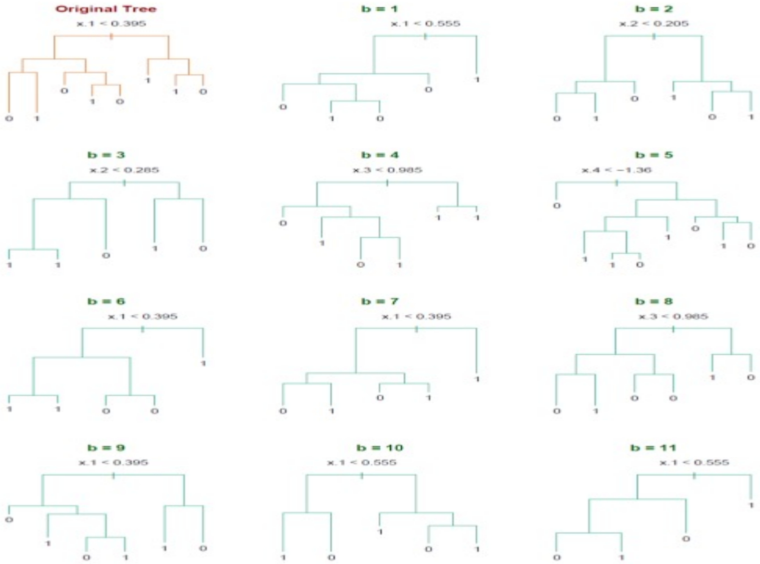
- A bootstrap sample is a smaller sample that is generated (bootstrapped) from a larger sample.
- It uses a resampling method with replacement.
- Bootstrap in many cases can result in less variance and more accurate results.



Bagging

- Bootstrap aggregation or bagging
  - General-purpose procedure for reducing the variance of a statistical learning method.
    - Given a set of n independent estimates $z_1, z_2, \ldots, z_n$ each with variance σ^2, the variance of their mean is n−times lower (σ^2/n).
      - When the estimates are not independent, reduction in variance is lower.
  - Uses multiple classifiers trained on different under-sampled subsets and then allow these classifiers to vote on a final decision.





Random forest algorithm

- Random forest classifier
  - Creates a set of decision trees from randomly selected subset of training set.
    - Each tree is built from a bootstrap sample of data
    - Form the tree based on the best feature from the subset
    - Repeat these steps T times, where T is the number of the trees
    - Impact on bias
      - Increases. Why?
        - Uses subsets of features in different independent trees
  - Aggregates the votes from different decision trees to decide the final class of the test object.

Random forest algorithm...

- In random forest:
  - all trees are fully grown and no pruning
  - we are dealing with two parameters:
    - number of trees (T);
      - Increasing the number can result in overfitting problem.
    - number of features in subtree (m_{try})

$$m_{try} = \sqrt{Number\ of\ features}$$

Random forest algorithm… Training

- For each of T iterations (T is the number of trees you may like to build):
  - Select a new bootstrap sample from the training set
  - Build an un-pruned tree on this bootstrap sample
  - At each internal node of the tree, randomly select m_{try} features and determine the best split using only these features.
    - Increasing number of features (m_{try}) for each split:
      - Increases correlation
      - Increases strength of single trees

Random forest algorithm… Testing

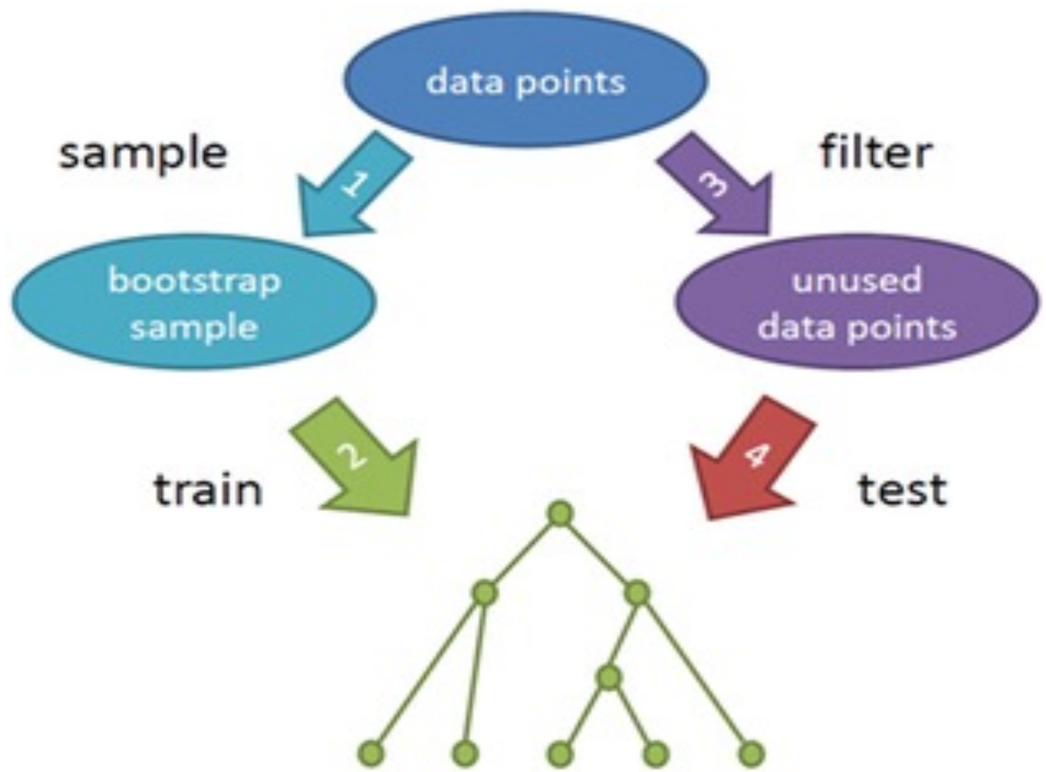
- Output overall prediction as a mean (or majority vote) from all individually trained trees.
- In random forest, the error rate depends on:
  - Correlation between trees (lower is better)
  - Strength of single trees (higher is better)

Out of bag error

- Estimate the goodness of fit of a bagged model
  - Out of Bag has been introduced which is equivalent to validation or test data.
- Each tree in a random forest is trained on a bootstrapped sample.
  - On average, each bagged tree makes use of 2/3 of the training instances.
- The remaining 1/3 of the instances are referred to as the out-of-bag (OOB) instances.

Out of bag error

- We can predict the response for the i^{th} observation using each of the trees in which that observation was OOB.
  - Average them to find the out of bag error



Feature importance

- Each node in the tree (single feature-based split)
  - How much each feature decreases the weighted impurity (Gini or Entropy) of the tree?
  - This provides rank of all features used in a tree.
- In Random Forest
  - Multiple trees
    - Multiple rank values of single feature
  - Average impurity decreasing scores across all trees for getting overall score and rank of the feature

Advantages/Disadvantages of Random Forest

- Fast to build and even faster to predict!
  - Fully parallelizable since you can parallelly run the trees to go even faster!
  - Decision Tree complexity is $O(d \times n \times logn)$.
  - A random forest with T trees would have $O(T \times d \times n \times logn)$, where d is the number of features and n is the number of data points
- Ability to handle data without pre-processing
  - Not always required to normalize your dataset before running this method
  - data does not need to be rescaled, transformed, or modified! (Resistant to outliers)
  - automatic handling of missing values (a property of decision trees)
- Less interpretable results than a single decision tree.

Ensemble Learning…

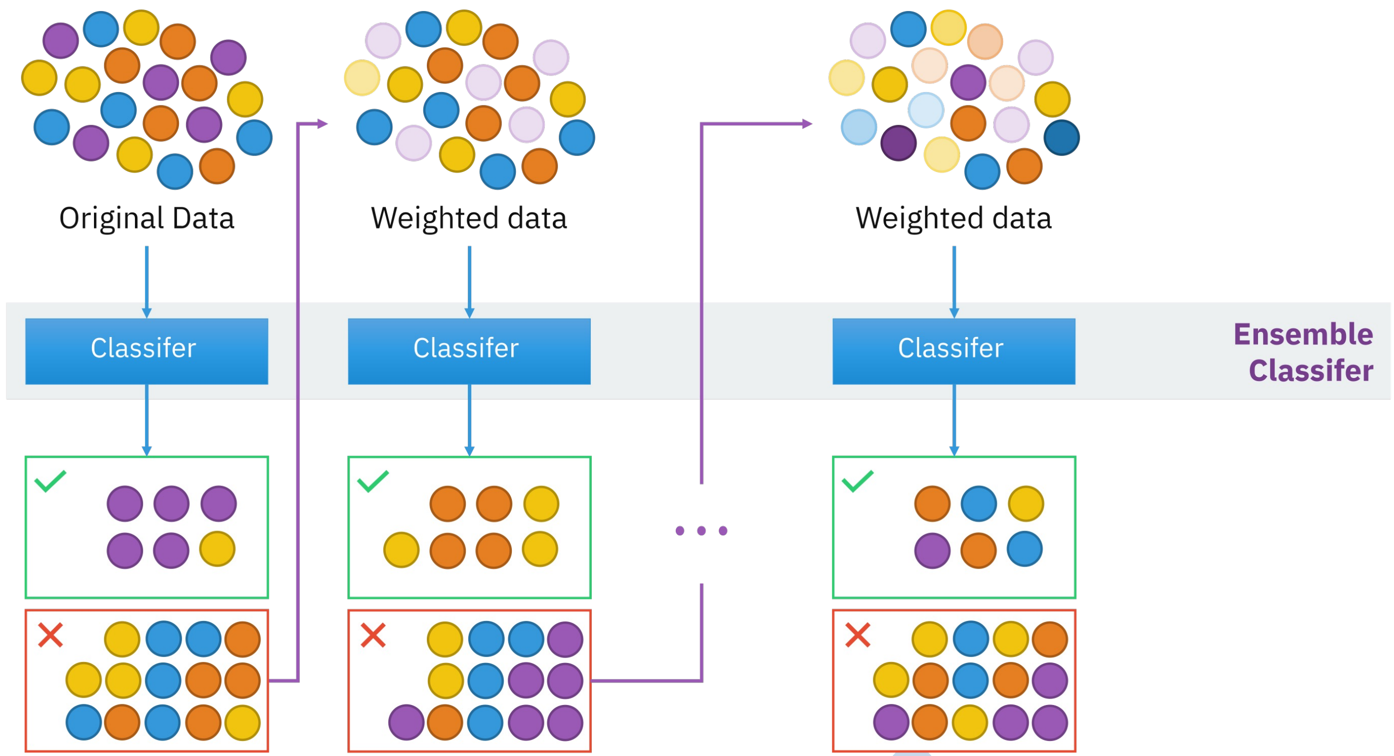
AdaBoost

Gradient Boost

XGBoost (Extreme Gradient Boosting)

Boosting

Training a bunch of individual models in a sequential way. Each individual model learns from mistakes made by the previous model.



Feature importance of using Random forest (RF)

- The significance of each feature in the input dataset can also be determined using Random Forest.
- Based on how much it helps to reduce impurity in the decision trees

References:

- Kursa, Miron B., and Witold R. Rudnicki. "The all relevant feature selection using random forest." arXiv preprint arXiv:1106.5112 (2011).
- Hasan, Md Al Mehedi, et al. "Feature selection for intrusion detection using random forest." Journal of information security 7.3 (2016): 129-140.
- Huljanah, Mia, et al. "Feature selection using random forest classifier for predicting prostate cancer." IOP Conference Series: Materials Science and Engineering. Vol. 546. No. 5. IOP Publishing, 2019.

Voting Classifier:

- An ensemble learning technique called a voting classifier combines the predictions of various separate classifiers to provide a final prediction.
- Several types of classifiers, such as Decision Trees, K-Nearest Neighbors, or Support Vector Machines, can be used individually.
- Each classifier is given one vote, and the final forecast is determined by a majority vote.
- Voting Classifier incorporates the benefits of various models while minimising the effects of their particular flaws.

Advance topic for SIT720

Stack Classifier:

- Another ensemble learning technique that aggregates the predictions of various separate classifiers is the Stack Classifier
- The first layer of a stack classifier comprises multiple separate classifiers that create predictions based on the input data.
- The second layer then integrates the previous layer's predictions to arrive at a final prediction.
- Several algorithms might be used at the second layer, including Decision Trees and Logistic Regression.
- Stack Classifier can increase the prediction's accuracy and generalizability by learning a more complicated decision boundary and minimising the chance of overfitting.

References:

1. Wolpert, David H. "Stacked generalization." Neural networks 5.2 (1992): 241-259.
2. Breiman, Leo. "Stacked regressions." Machine learning 24 (1996): 49-64.
3. Yousaf, Anam, et al. "Emotion recognition by textual tweets classification using voting classifier (LR-SGD)." IEEE Access 9 (2020): 6286-6295.
4. Khan, Muhammad Almas, et al. "Voting classifier-based intrusion detection for iot networks." Advances on Smart and Soft Computing: Proceedings of ICACIn 2021. Springer Singapore, 2022.

Thank You.