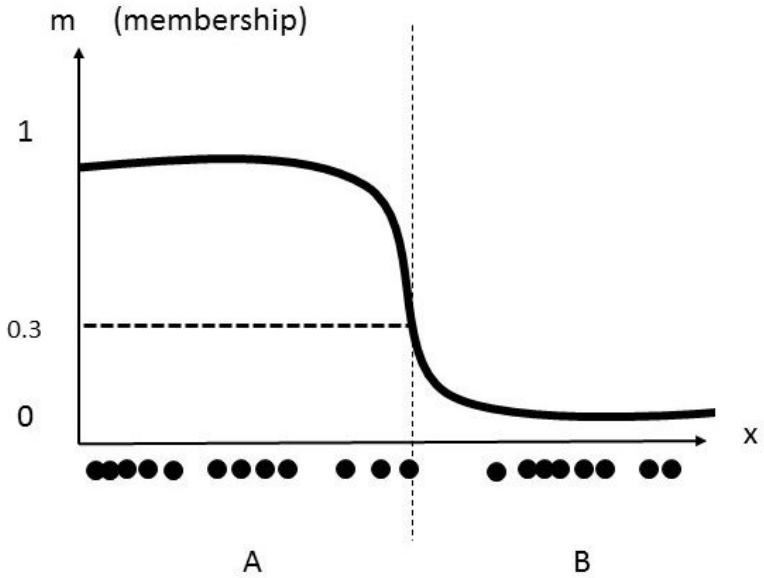
Fuzzy Clustering Ensemble (Boosting)

Some Ideas

C-Means

- An extension of K-Means
- In fuzzy clustering, data points can potentially belong to multiple clusters.



FCM - Algorithm

The fuzzy c-means algorithm is very similar to the k-means algorithm:

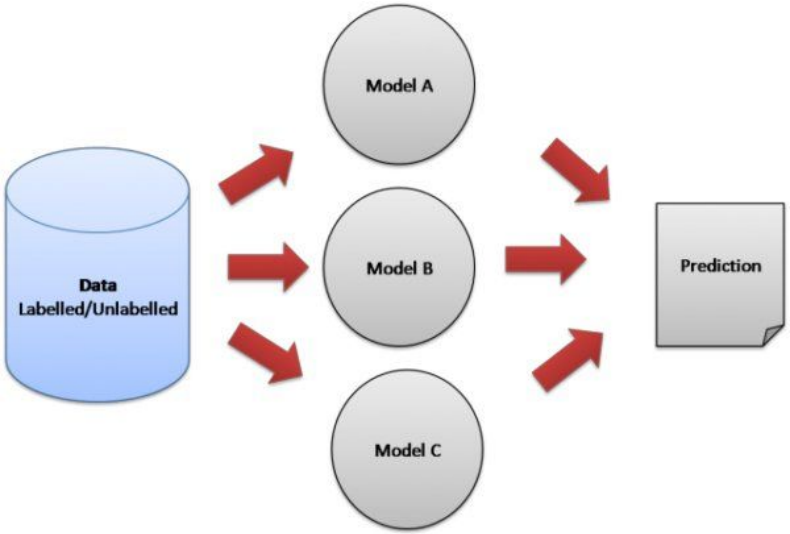
- Choose a number of clusters.
- Assign coefficients randomly to each data point for being in the clusters.
- Repeat until the algorithm has converged (that is, the coefficients' change between two iterations is no more than epsilon, the given sensitivity threshold) :
  - Compute the centroid for each cluster (shown below).
  - For each data point, compute its coefficients of being in the clusters.

C-Means - Minimize Cost Function

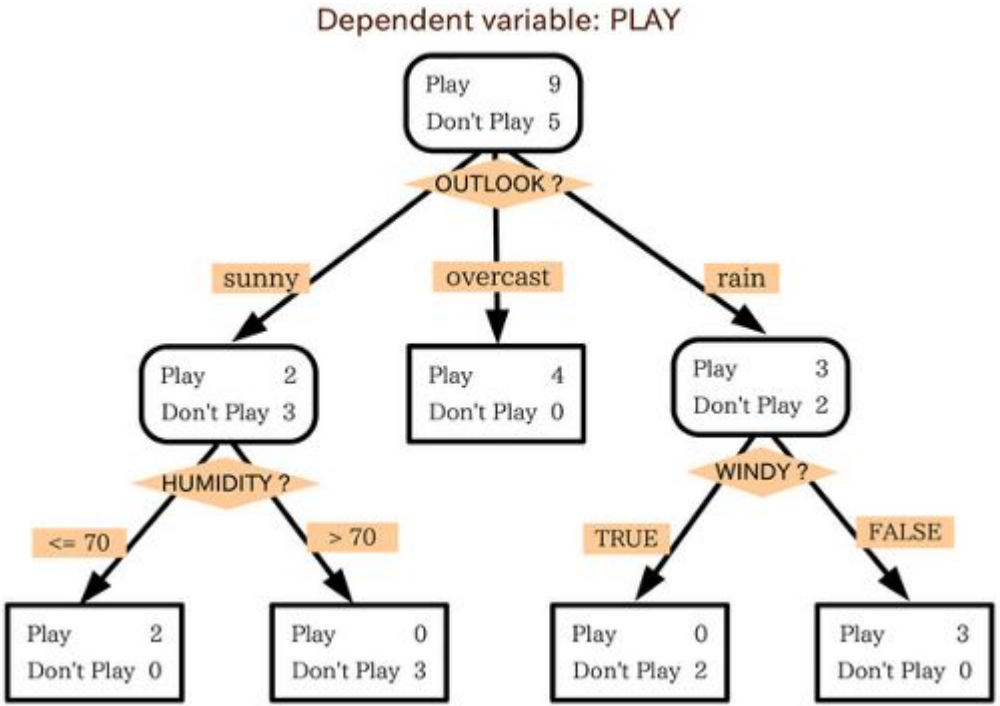
$$\arg\min_{C} \sum_{i=1}^{n} \sum_{j=1}^{c} w_{ij}^{m} \left\| \mathbf{x}_i - \mathbf{c}_j \right\|^2,$$

Ensemble Models

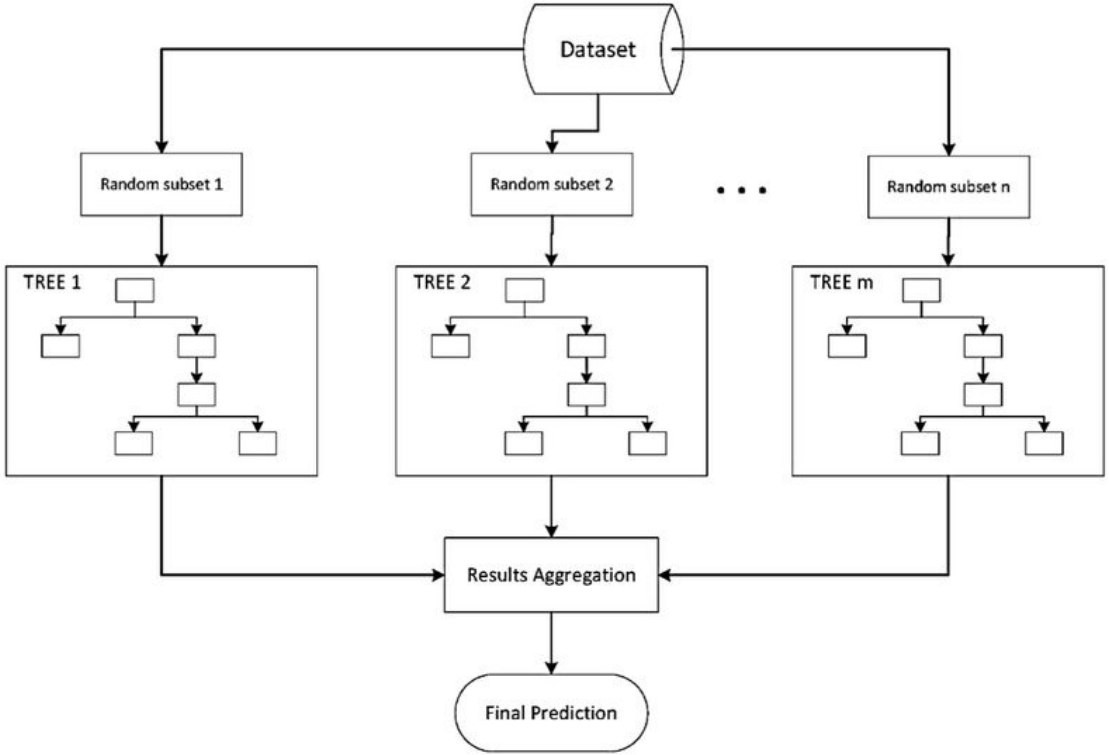
Ensemble methods is a machine learning technique that combines several base models in order to produce one optimal predictive model.



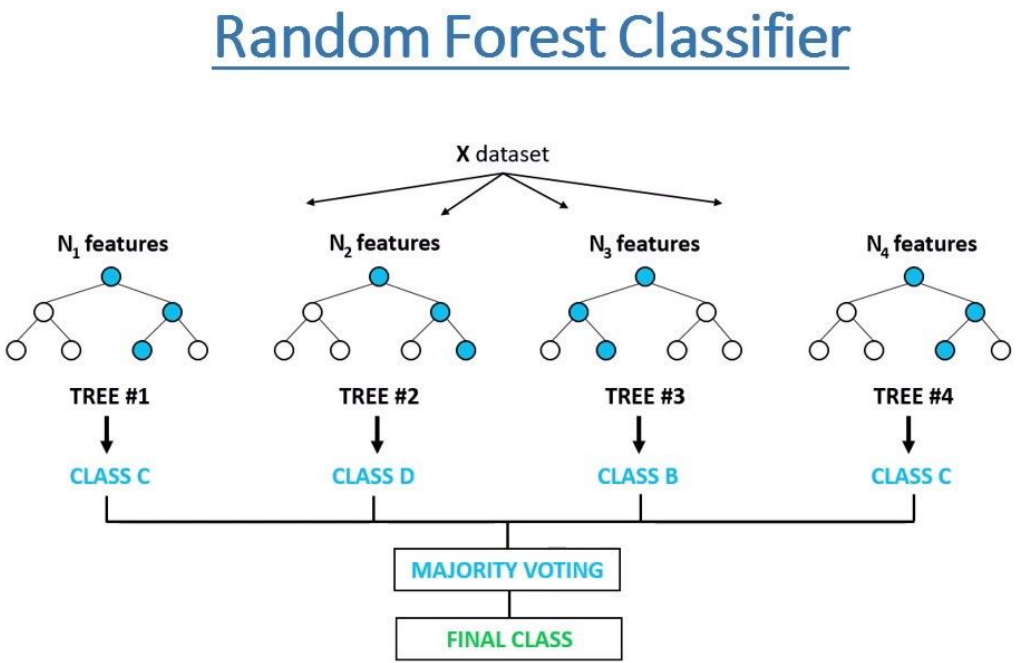
Decision Tree



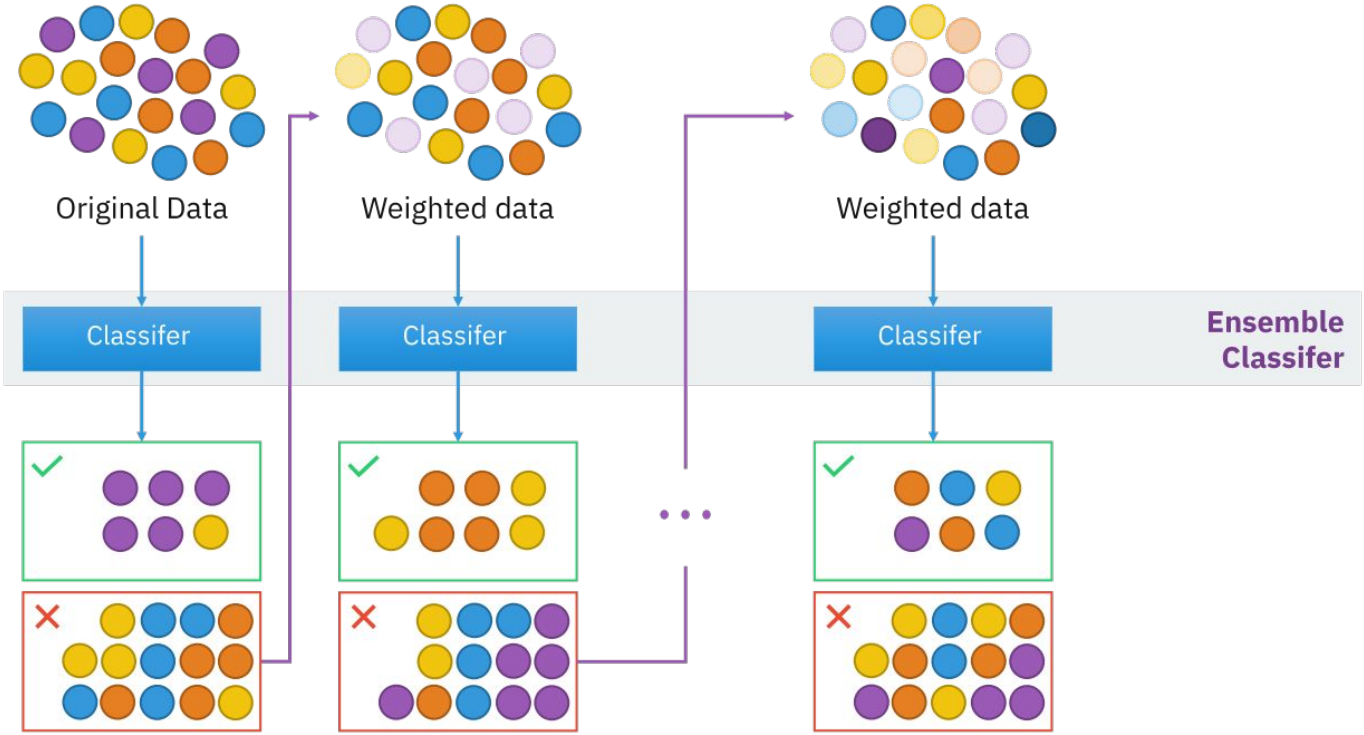
Bagging - ***B***ootstrap ***AGG***regating



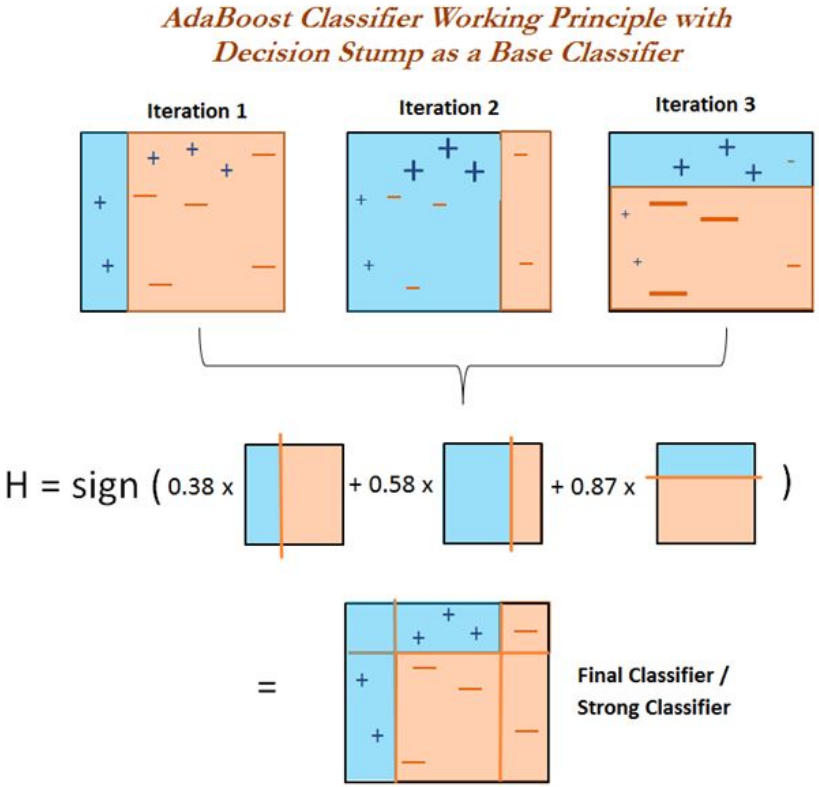
Random Forest



Boosting - Weak Classifiers



Boosting


AdaBoost Classifier Working Principle with Decision Stump as a Base Classifier
Iteration 1
Iteration 2
Iteration 3
H = sign (0.38 x
+ 0.58 x
+ 0.87 x
)
=
Final Classifier / Strong Classifier