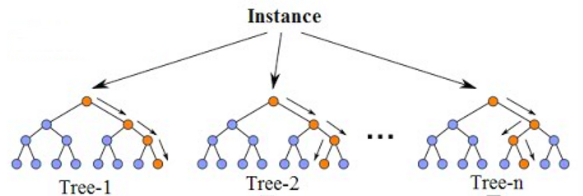
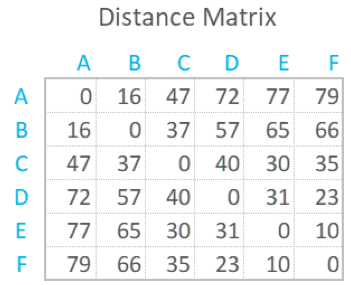
**DAY-11 [14-10-2021] Random Forest, Regression Trees:**

**Random Forest:** A random forest is a supervised machine learning algorithm that is constructed from decision tree algorithms. This algorithm is applied in various industries such as banking and e-commerce to predict behavior and outcomes.



**Proximity matrix**: Profile matrices contain measurements of variables for a set of items. Proximity matrices contain measurements of relations, or proximities, between items. If the measurements in a proximity matrix tell how close things are to each other then you have a similarity matrix.

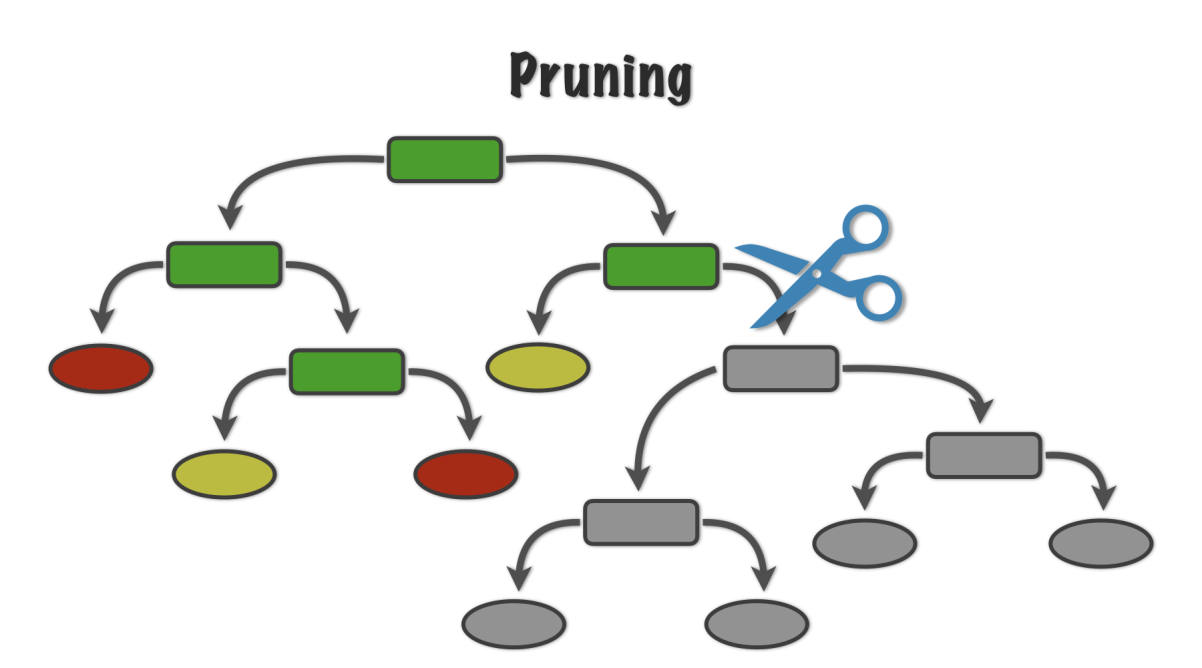
**Distance Matrix**: In mathematics, computer science and especially graph theory, a distance matrix is a square matrix containing the distances, taken pairwise, between the elements of a set. Depending upon the application involved, the distance being used to define this matrix may or may not be a metric.



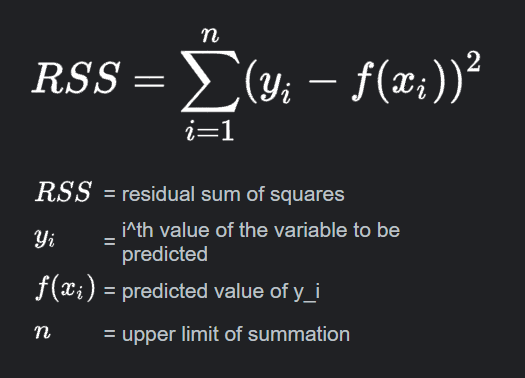
**Bootstrapping**: Bootstrapping is any test or metric that uses random sampling with replacement, and falls under the broader class of resampling methods. Bootstrapping assigns measures of accuracy to sample estimates. This technique allows estimation of the sampling distribution of almost any statistic using random sampling methods.

**Out-of-Bag Error**: Out-of-bag error, also called out-of-bag estimate, is a method of measuring the prediction error of random forests, boosted decision trees, and other machine learning models utilizing bootstrap aggregating. Bagging uses subsampling with replacement to create training samples for the model to learn from.

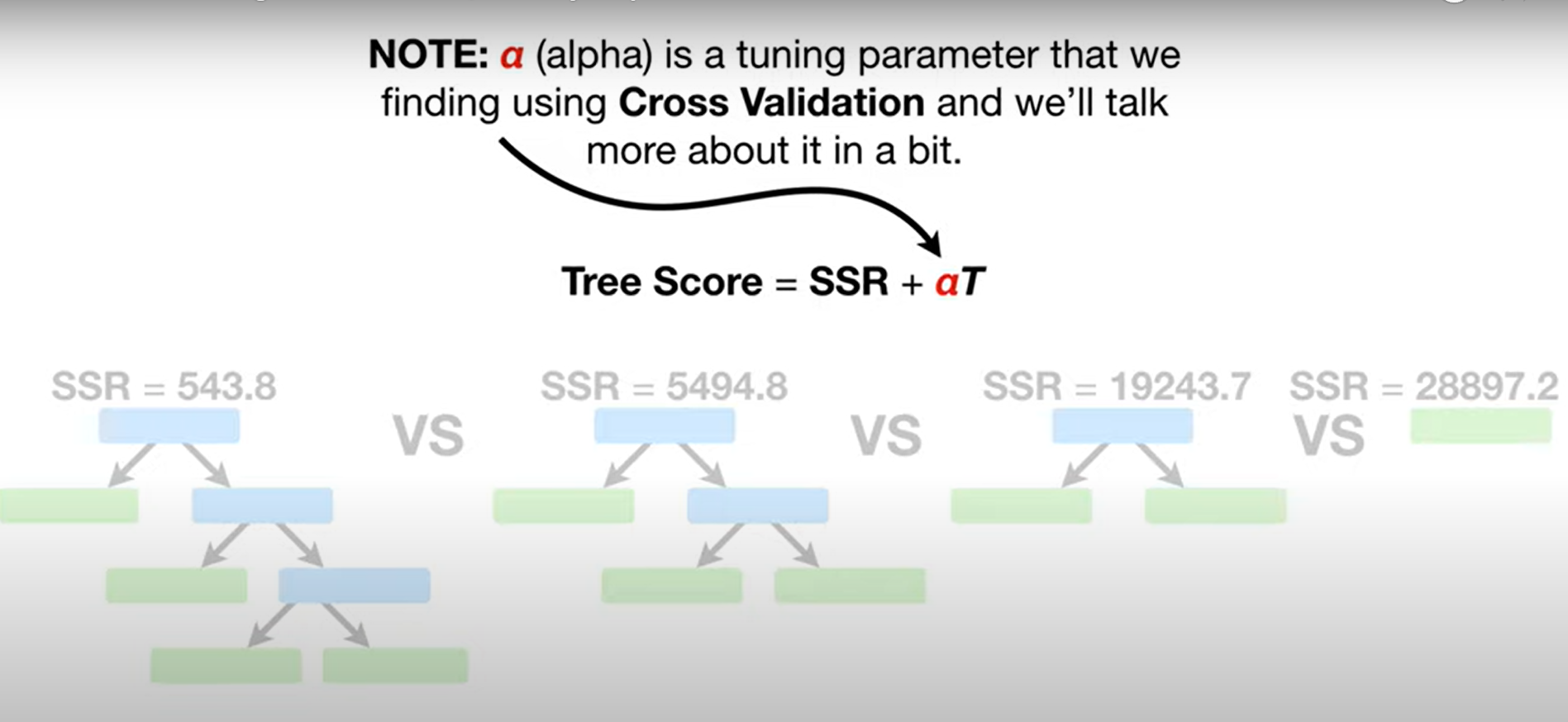
**Pruning of Decision trees**: Pruning is a data compression technique in machine learning and search algorithms that reduces the size of decision trees by removing sections of the tree that are non-critical and redundant to classify instances.

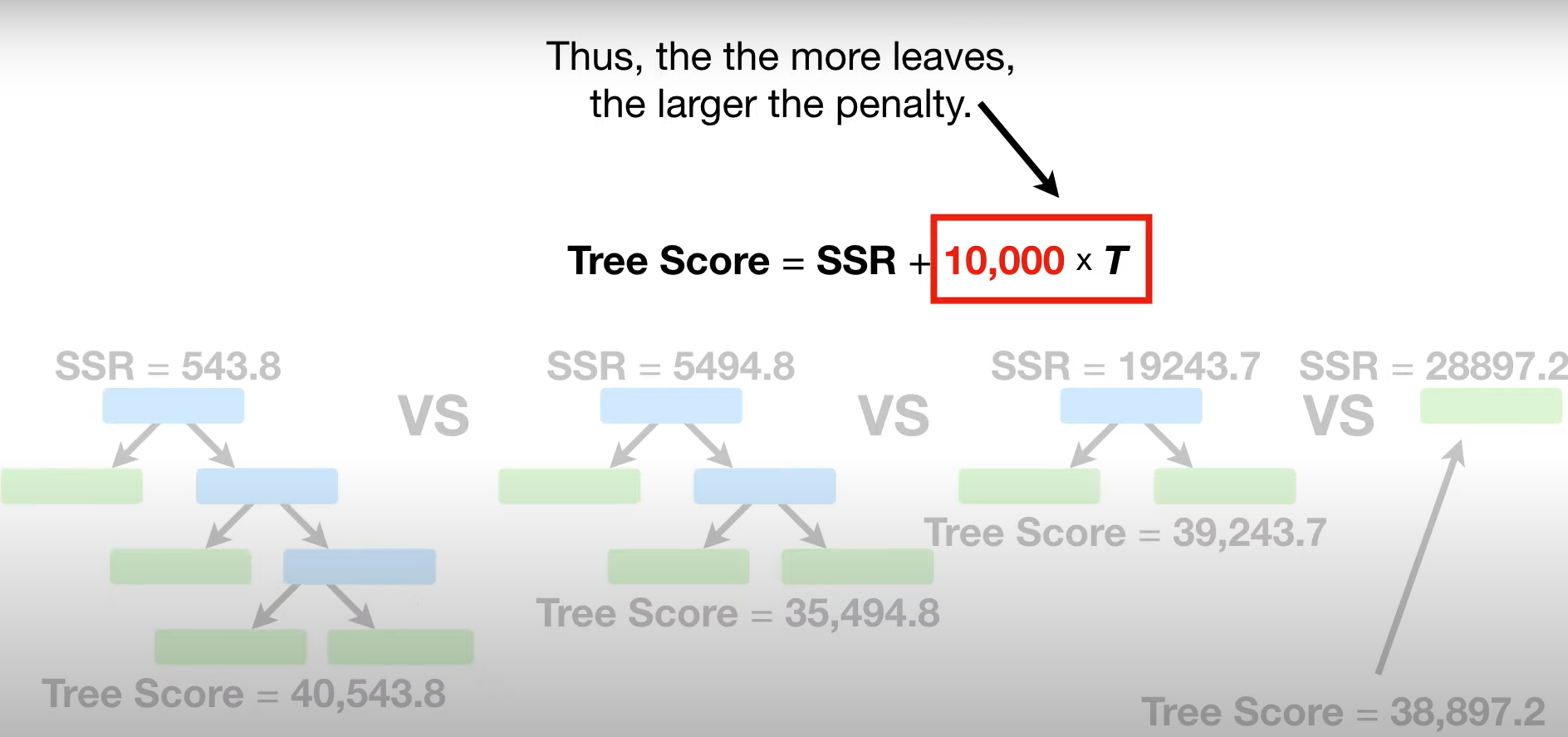


**Residual sum of squares for Pruning Decision trees**: In statistics, the residual sum of squares, also known as the sum of squared residuals or the sum of squared estimate of errors, is the sum of the squares of residuals. It is a measure of the discrepancy between the data and an estimation model, such as a linear regression.



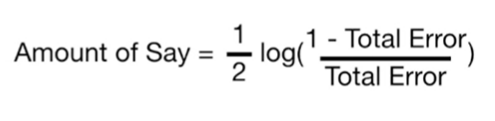
**Tree Score**: Used for pruning

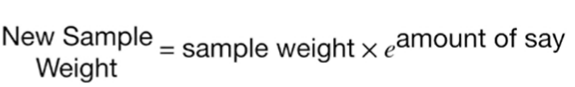


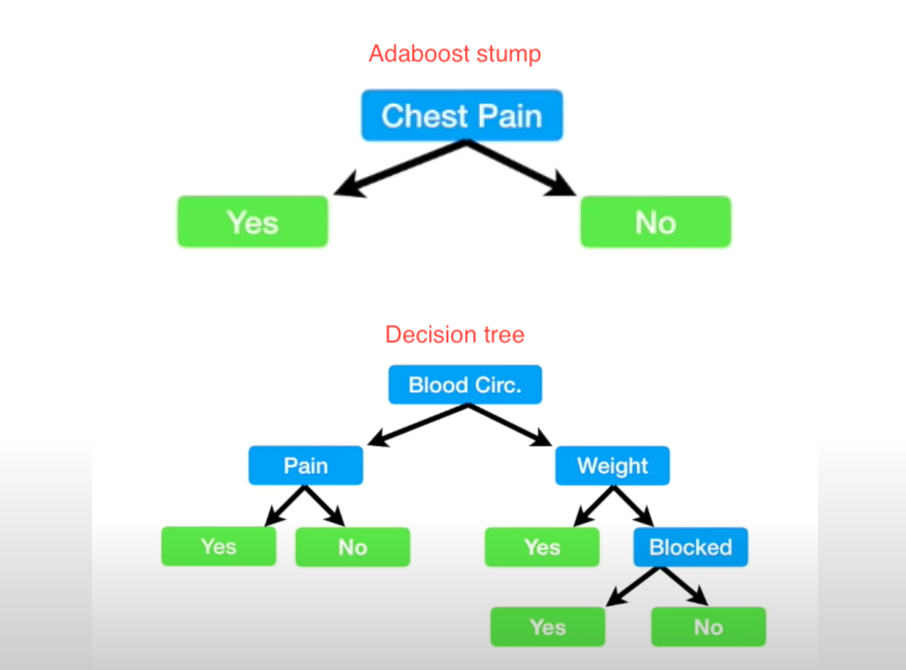


**Adaboost Classifier**: An AdaBoost [1] classifier is a meta-estimator that begins by fitting a classifier on the original dataset and then fits additional copies of the classifier on the same dataset but where the weights of incorrectly classified instances are adjusted such that subsequent classifiers focus more on difficult cases.

Amount of Say and Sample weight:







**References:**

Random forest part-1: <https://www.youtube.com/watch?v=J4Wdy0Wc_xQ&t=240s>

Random forest part-2: <https://www.youtube.com/watch?v=sQ870aTKqiM&t=589s>

Pruning regression trees: <https://www.youtube.com/watch?v=D0efHEJsfHo>

Adaboost: <https://www.youtube.com/watch?v=LsK-xG1cLYA>

**Other sites:**

1. <https://www.datacamp.com/community/tutorials/decision-tree-classification-python?utm_source=adwords_ppc&utm_campaignid=1455363063&utm_adgroupid=65083631748&utm_device=c&utm_keyword=&utm_matchtype=b&utm_network=g&utm_adpostion=&utm_creative=332602034358&utm_targetid=aud-299261629574:dsa-429603003980&utm_loc_interest_ms=&utm_loc_physical_ms=9062011&gclid=Cj0KCQjwqp-LBhDQARIsAO0a6aJxN7t3yiJBHSnr-eVbnxX9f7VdXLDV6RR2L_MJqBV_q081rOmOp6gaAs2WEALw_wcB>
2. <https://www.analyticsvidhya.com/blog/2020/10/all-about-decision-tree-from-scratch-with-python-implementation/>
3. <https://towardsdatascience.com/decision-trees-a-step-by-step-approach-to-building-dts-58f8a3e82596>