

## Classification 2

### ➤ Decision Tree

- *Node*: A point in the decision tree where a decision is made or where the tree branches
- *Leaf node*: A leaf node is a terminal node in a decision tree, where the impurity will be 0
- *Root node*: It is the topmost node in the decision tree
- *Split*: The division of a node into 2 based on a specific condition.
- *Decision Stumps*: Small, shallow trees with only one split.
- *Categorical vs Continues*
- *Random Split*: Instead of choosing the best split, select feature and threshold randomly
- *Recursive binary split*: Repeatedly splitting nodes into two child nodes until a stopping criterion is met
- *Entropy*: a measure of impurity in a set of data
- *Information Gain*: measures the reduction in entropy achieved by a particular split
- *Gini Impurity*: Another measure of impurity
- *Pruning*: the process of removing parts of the tree that do not provide significant power in predicting target values
- *Pruning techniques* (Reduced error pruning, Cost complexity)
- *Overfitting*: occurs when the model is too complex
- *Decision Boundary*: border that separates different classes in the feature space.
- *Recursive Partitioning*: The process of repeatedly dividing the dataset into subsets based on specific condition
- *Variance Reduction*
- *CART algorithm*: Uses gini impurity
- *C4.5 Algorithm*: uses information gain
- *Chi-squared test*: A statistical test used in a decision tree algorithm
- *CHAID* (Chi-squared Automatic Interaction Detection)
- *Termination criteria*: Condition that determines when to stop growing the tree
- *Parameters*
- *Criterion*: Splitting criterion (e.g., 'gini', 'entropy', 'mse').
- *Max Depth*: Maximum depth of the tree.
- *Min Samples Split*: Minimum samples required to split a node.
- *Min Samples Leaf*: Minimum samples required in a leaf node.
- *Max Features*: Maximum number of features to consider for a split.
- *Presort*: Whether to presort the data.
- *Random State*: Seed for random number generator.

### ➤ Random Forest

- *Ensemble Learning*: Use multiple models and combine all of their outputs
- *Bagging*: Training multiple instances of the same algorithm
- *Bootstrapped Samples*: random samples drawn with replacement from the original dataset

- *Feature Subsetting: Involves selecting a random set for each step of training*
- *Voting: Used in ensemble learning*
- *Out-of-Bag Error: The error rate of the model on the samples that were not included in its training set*
- *Randomness*
- *Feature Importance: measures the contribution of each feature to the performance of the model*
- *Bias-Variance Trade-off*
- *Random Subspace Method*
- *Random Forest Variants (e.g., Extremely Randomized Trees)*
- *Feature Importances (Gini Importance, Mean Decrease Accuracy)*
- *Feature Permutation Importance*
- *Bootstrapping*
- *Proximity Matrix: measure how often 2 samples end up in the same leaf node*
- *Parallel Processing: Many ensemble methods can leverage parallel processing to speed up training*
- *Random Subspace Method*
- *Stochastic Random Forest*
- *Unbiasedness of Random forest*
- *Permutation Feature Importance*
- *Conditional Variable Importance*
- *Model Interpretability*

#### *Parameters*

- *Number of Estimators: Number of decision trees in the forest.*
- *Max Features: Maximum number of features to consider for a split.*
- *Criterion: Splitting criterion.*
- *Max Depth: Maximum depth of each tree.*
- *Min Samples Split: Minimum samples required to split a node.*
- *Min Samples Leaf: Minimum samples required in a leaf node.*
- *Bootstrap: Whether to use bootstrapping.*
- *Random State: Seed for random number generator.*

#### ➤ *SVM*

- *Hyperplane: Decision boundary that separates data points belonging to different classes in feature space*
- *Margin: distance between the plane hyperplane and nearest data point*
- *Support Vectors: Data points that lie closest to the decision boundary (hyperplane)*
- *Kernel Function (Linear, Polynomial, RBF):*
- *Dual Problem*
- *C parameter*
- *Hard Margin, Soft Margin*
- *Multiclass SVM*
- *Hinge Loss*
- *Kernel Trick*
- *Kernel Matrix*
- *Mercer's Theorem*
- *Slack Variables (C-SVM)*

- *Non-Linear Separability*
- *Margin Violation*
- *Sequential Minimal Optimization (SMO)*
- *Twin SVM*
- *Multi-class SVM*
- *Nu-SVM*
- *Margin Maximization*

#### *Parameters*

- *C (Regularization Parameter): Controls the trade-off between maximizing the margin and minimizing the classification error.*
- *Kernel: Kernel function (e.g., 'linear', 'poly', 'rbf').*
- *Degree (for 'poly' kernel): Degree of the polynomial kernel.*
- *Gamma (Kernel Coefficient): Defines the shape of the kernel.*
- *Coef0 (for 'poly' and 'sigmoid' kernels): Independent term in the kernel function.*
- *Class Weight: Optional weights for classes.*
- *Decision Function Shape: Decision function shape ('ovo' or 'over') for multiclass classification.*
- *Probability: Whether to enable probability estimates.*