Assignment 8
DECISIONTREECLASSIFIERAPI

Arham Lodha J031

• class sklearn.tree.DecisionTreeClassifier(*, criterion='gini', splitter='best',max_depth=None, min_samples_split=2, min_samples_leaf=1, min_weight_fraction_leaf=0.0, max_features=None, random_state=None, max_leaf_nodes=None, min_impurity_decrease=0.0, min_impurity_split=None,class_weight=None,ccp_alpha=0.0)

• Parameters–

1. criterion:{“gini”,“entropy”},default=”gini”
2. splitter:{“best”,“random”},default=”best”
3. max_depth:int,default=None
4. min_samples_split:intorfloat,default=2
5. min_samples_leaf:intorfloat,default=1
6. min_weight_fraction_leaf:float,default=0.0
7. max_features:int,floator{“auto”,“sqrt”,“log2”},default=None
8. random_state:int,RandomStateinstanceorNone,default=None
9. max_leaf_nodes: int, default=None10.min_impurity_decrease: float, default=0.011.min_impurity_split:float,defau lt=0
12. class_weight:dict,listofdictor“balanced”,default=None
13. ccp_alpha:non-negativefloat,default=0.0

• Attributes-

1. classes_ndarrayofshape(n_classes,)orlistofndarray
2. feature_importances_ndarrayofshape(n_features,)
3. max_features_int
4. n_classes_intorlistofint
5. n_features_int
6. n_outputs_int
7. tree_Treeinstance

- **Advantages:-**

1. Simpletounderstandandtointerpret.Treescanbevisualised.

2. Requires littledatapreparation.

3. The cost of using the tree (i.e., predicting data) is logarithmic in thenumberof datapoints usedto train thetree.

4. Abletohandlebothnumericalandcategoricaldata.

5. Abletohandlemulti-outputproblems.

6. Possible to validate a model using statistical tests. That makes itpossibleto accountforthereliabilityof themodel.

7. Performs well even if its assumptions are somewhat violated by thetruemodelfromwhich thedataweregenerated.

- **Disadvantages:**

1. Decision-tree learners can create over-complex trees that do notgeneralisethedatawell. Thisis called overfitting.

2. Decisiontreescanbeunstablebecausesmallvariationsinthedatamigh t result in a completely different tree being generated. Thisproblemismitigatedbyusingdecision treeswithinan ensemble.

3. Predictions of decision trees are neither smooth nor continuous, butpiecewise constant approximations as seen in the above figure.Therefore,theyarenotgoodatextrapolation.

4. Decision tree learners create biased trees if some classes dominate. Itisthereforerecommended to balancethedatasetprior tofittingwith

thedecisiontree.