Christopher Morgan

CS478: Brother Christophe

Project 1: Decision Tree

**Correctly implement the ID3**

* + **An option for choosing the splitting criterion: information gain or accuracy.**

The command line argument –UA (Use Accuracy) will flip the algorithm from its default Entropy minimization to maximizing Accuracy.

* + **Some mechanism to handle continuous-valued attributes.**

My algorithm will simply divide a continuous-valued attribute in two. The attribute is split along the mean of its data values.

* + **Some mechanism to handle unknown (or missing) attribute values.**

My algorithm uses a preprocessing technique to fill in any missing attributes. The preprocessing strategy replaces any missing attribute values with the majority value for that attribute. This seems reasonable given that majority is the best guess we can make about a given attribute value.

* + **You should be able to get ~68% predictive accuracy on lenses data with cross-validation.**

|  |  |
| --- | --- |
| Dataset name: ../data/lenses.arff  Number of instances (rows): 24  Number of attributes (cols): 5  Learning algorithm: DecisionTree  Evaluation method: cross  Rep: 0, Fold: 0, Accuracy: 0.5  Rep: 0, Fold: 1, Accuracy: 0  Rep: 0, Fold: 2, Accuracy: 1  Rep: 0, Fold: 3, Accuracy: 1  Rep: 0, Fold: 4, Accuracy: 1  Rep: 0, Fold: 5, Accuracy: 0.5  Rep: 0, Fold: 6, Accuracy: 1  Rep: 0, Fold: 7, Accuracy: 0.5  Rep: 0, Fold: 8, Accuracy: 1  Rep: 0, Fold: 9, Accuracy: 0.5  Mean predictive accuracy: **0.7** | **Using Accuracy:**  Dataset name: ../data/lenses.arff  Number of instances (rows): 24  Number of attributes (cols): 5  Learning algorithm: DecisionTree  Evaluation method: cross  Rep: 0, Fold: 0, Accuracy: 1  Rep: 0, Fold: 1, Accuracy: 0.5  Rep: 0, Fold: 2, Accuracy: 1  Rep: 0, Fold: 3, Accuracy: 0.5  Rep: 0, Fold: 4, Accuracy: 0  Rep: 0, Fold: 5, Accuracy: 0.5  Rep: 0, Fold: 6, Accuracy: 1  Rep: 0, Fold: 7, Accuracy: 1  Rep: 0, Fold: 8, Accuracy: 0.5  Rep: 0, Fold: 9, Accuracy: 1  Mean predictive accuracy: **0.7** |

**Use your ID3 algorithm on the**[**Iris**](http://dml.cs.byu.edu/~cgc/docs/mldm_tools/Assignments/Datasets/iris.arff)**problem.**

* + **Induce a decision tree using the entire dataset with information gain as the splitting criterion. Give a visual representation of the tree.**

\* See “1) Iris: Entropy” below.

* + **Induce a decision tree using the entire dataset with accuracy as the splitting criterion. Give a visual representation of the tree. Compare this tree with the one obtained with information gain as the splitting criterion.**

See “2) Iris: Accuracy” below.

The main order of the Entropy algorithm was to choose petalLength, petalWidth, and then sepalWidth in that order. However, when switching to the Accuracy algorithm it favored petalWidth, petalLenght, then sepalLength in that order. Changing algorithms to accuracy therefore cause an increase in weight for petalWidth, cause it to be selected first over petalLength. In addition to the order of attributes chosen the switch in algorithms also created a slightly larger tree when accuracy was used. The Entropy tree consisted of 19 nodes where the Accuracy tree was made up of 21.

* + Evaluate predictive accuracy using 10-fold cross-validation for information gain and accuracy. Compare the results.

**1) Iris: Entropy**

Node: Property: petallength

Node: Property: petalwidth

Node: Class: 0

Node: Property: sepalwidth

Node: Class: 1

Node: Class: 0

Node: Property: petalwidth

Node: Property: sepalwidth

Node: Property: sepallength

Node: Class: 1

Node: Class: 1

Node: Property: sepallength

Node: Class: 1

Node: Class: 1

Node: Property: sepalwidth

Node: Class: 2

Node: Property: sepallength

Node: Class: 2

Node: Class: 2

**2) Iris: Accuracy**

Node: Property: petalwidth

Node: Property: petallength

Node: Class: 0

Node: Property: sepallength

Node: Property: sepalwidth

Node: Class: 1

Node: Class: 0

Node: Class: 1

Node: Property: petallength

Node: Property: sepallength

Node: Property: sepalwidth

Node: Class: 1

Node: Class: 1

Node: Property: sepalwidth

Node: Class: 1

Node: Class: 1

Node: Property: sepallength

Node: Property: sepalwidth

Node: Class: 2

Node: Class: 2

Node: Class: 2

**3) Voting :: Entropy**

Node: Property: 'physician-fee-freeze'

Node: Property: 'synfuels-corporation-cutback'

Node: Property: 'crime'

Node: Property: 'anti-satellite-test-ban'

Node: Property: 'education-spending'

Node: Class: 0

Node: Class: 1

Node: Class: 0

Node: Property: 'anti-satellite-test-ban'

Node: Class: 0

Node: Property: 'el-salvador-aid'

Node: Property: 'adoption-of-the-budget-resolution'

Node: Property: 'handicapped-infants'

Node: Class: 0

Node: Property: 'water-project-cost-sharing'

Node: Property: 'religious-groups-in-schools'

Node: Class: 1

Node: Class: 0

NULL

Node: Class: 0

Node: Property: 'handicapped-infants'

Node: Property: 'religious-groups-in-schools'

Node: Class: 0

Node: Property: 'education-spending'

Node: Property: 'water-project-cost-sharing'

Node: Class: 1

Node: Property: 'adoption-of-the-budget-resolution'

NULL

Node: Property: 'aid-to-nicaraguan-contras'

NULL

Node: Property: 'mx-missile'

NULL

Node: Property: 'immigration'

NULL

Node: Property: 'superfund-right-to-sue'

NULL

Node: Property: 'duty-free-exports'

Node: Property: 'export-administration-act-south…

Node: Class: 0

Node: Class: 1

NULL

Node: Class: 0

Node: Class: 0

Node: Class: 0

Node: Property: 'synfuels-corporation-cutback'

Node: Property: 'duty-free-exports'

Node: Property: 'adoption-of-the-budget-resolution'

Node: Class: 1

Node: Property: 'anti-satellite-test-ban'

Node: Property: 'handicapped-infants'

Node: Property: 'immigration'

Node: Class: 0

Node: Class: 1

NULL

Node: Class: 1

Node: Property: 'immigration'

Node: Property: 'handicapped-infants'

Node: Property: 'anti-satellite-test-ban'

Node: Property: 'education-spending'

Node: Class: 0

Node: Class: 1

Node: Class: 1

NULL

Node: Class: 1

Node: Property: 'adoption-of-the-budget-resolution'

Node: Property: 'el-salvador-aid'

Node: Class: 0

Node: Property: 'immigration'

Node: Property: 'superfund-right-to-sue'

Node: Class: 0

Node: Property: 'anti-satellite-test-ban'

Node: Property: 'aid-to-nicaraguan-contras'

Node: Property: 'water-project-cost-sharing'

Node: Property: 'handicapped-infants'

Node: Property: 'religious-groups-in-schools'

NULL

Node: Property: 'mx-missile'

Node: Property: 'education-spending'

NULL

Node: Property: 'crime'

NULL

Node: Property: 'duty-free-exports'

Node: Property: 'export-administration-act-south…

Node: Class: 0

Node: Class: 1

NULL

NULL

Node: Class: 1

Node: Property: 'handicapped-infants'

Node: Class: 1

Node: Property: 'religious-groups-in-schools'

NULL

Node: Property: 'mx-missile'

Node: Property: 'education-spending'

Node: Property: 'crime'

NULL

Node: Property: 'duty-free-exports'

Node: Property: 'export-administration-act-south…

Node: Class: 0

Node: Class: 0

NULL

Node: Class: 1

NULL

Node: Class: 1

Node: Class: 1

Node: Class: 1

Node: Property: 'anti-satellite-test-ban'

Node: Class: 0

Node: Class: 1

**4)Voting: Accuracy:**

Node: Property: 'physician-fee-freeze'

Node: Property: 'handicapped-infants'

Node: Property: 'religious-groups-in-schools'

Node: Class: 0

Node: Property: 'water-project-cost-sharing'

Node: Property: 'adoption-of-the-budget-resolution'

Node: Property: 'el-salvador-aid'

Node: Class: 0

Node: Property: 'education-spending'

Node: Class: 1

Node: Class: 0

Node: Class: 0

Node: Property: 'adoption-of-the-budget-resolution'

Node: Class: 0

Node: Property: 'mx-missile'

Node: Class: 0

Node: Property: 'el-salvador-aid'

Node: Property: 'anti-satellite-test-ban'

Node: Property: 'aid-to-nicaraguan-contras'

NULL

Node: Property: 'education-spending'

Node: Class: 0

Node: Class: 1

Node: Class: 0

Node: Property: 'anti-satellite-test-ban'

NULL

Node: Property: 'aid-to-nicaraguan-contras'

Node: Class: 0

Node: Property: 'immigration'

NULL

Node: Property: 'superfund-right-to-sue'

Node: Class: 0

Node: Property: 'synfuels-corporation-cutback'

Node: Property: 'education-spending'

Node: Property: 'crime'

NULL

Node: Property: 'duty-free-exports'

Node: Property: 'export-administration-act-south…

Node: Class: 0

Node: Class: 1

NULL

NULL

NULL

Node: Property: 'water-project-cost-sharing'

Node: Property: 'anti-satellite-test-ban'

Node: Class: 0

Node: Property: 'adoption-of-the-budget-resolution'

Node: Property: 'el-salvador-aid'

Node: Property: 'duty-free-exports'

Node: Class: 1

Node: Class: 0

Node: Class: 0

Node: Class: 0

Node: Class: 0

Node: Property: 'handicapped-infants'

Node: Property: 'water-project-cost-sharing'

Node: Property: 'adoption-of-the-budget-resolution'

Node: Property: 'religious-groups-in-schools'

Node: Class: 1

Node: Property: 'el-salvador-aid'

Node: Class: 1

Node: Property: 'anti-satellite-test-ban'

Node: Property: 'aid-to-nicaraguan-contras'

Node: Property: 'mx-missile'

Node: Property: 'synfuels-corporation-cutback'

Node: Class: 1

Node: Property: 'export-administration-act-south-africa'

Node: Class: 0

Node: Class: 1

NULL

NULL

Node: Class: 1

Node: Property: 'el-salvador-aid'

Node: Class: 1

Node: Property: 'religious-groups-in-schools'

NULL

Node: Property: 'aid-to-nicaraguan-contras'

Node: Property: 'education-spending'

Node: Class: 0

Node: Class: 1

Node: Class: 1

Node: Property: 'el-salvador-aid'

Node: Class: 0

Node: Property: 'adoption-of-the-budget-resolution'

Node: Property: 'superfund-right-to-sue'

Node: Class: 0

Node: Property: 'religious-groups-in-schools'

Node: Class: 1

Node: Property: 'mx-missile'

Node: Property: 'anti-satellite-test-ban'

Node: Property: 'duty-free-exports'

Node: Class: 1

Node: Class: 0

Node: Class: 1

Node: Class: 1

Node: Property: 'religious-groups-in-schools'

NULL

Node: Property: 'synfuels-corporation-cutback'

Node: Property: 'aid-to-nicaraguan-contras'

Node: Property: 'immigration'

Node: Class: 0

Node: Class: 1

Node: Class: 1

Node: Class: 0

Node: Property: 'synfuels-corporation-cutback'

Node: Class: 1

Node: Property: 'mx-missile'

Node: Property: 'water-project-cost-sharing'

Node: Class: 1

Node: Property: 'adoption-of-the-budget-resolution'

Node: Property: 'el-salvador-aid'

NULL

Node: Property: 'religious-groups-in-schools'

NULL

Node: Property: 'anti-satellite-test-ban'

Node: Property: 'aid-to-nicaraguan-contras'

Node: Property: 'immigration'

Node: Property: 'education-spending'

Node: Property: 'superfund-right-to-sue'

NULL

Node: Property: 'crime'

NULL

Node: Property: 'duty-free-exports'

Node: Property: 'export-administration-act-south…

Node: Class: 0

Node: Class: 0

NULL

Node: Class: 1

Node: Class: 1

Node: Class: 1

NULL

Node: Class: 0

Node: Property: 'water-project-cost-sharing'

NULL

Node: Property: 'immigration'

Node: Class: 0

Node: Class: 1