Mimir

Search

Daniel
Engbert

Home   CMSC 471 Sp17   Project 3: Decision Trees

Project 3: Decision Trees   *Active*   Starter Code   Open Workspace   **Submit Files**

Project 3: Decision Trees

You are going to implement the ID3 algorithm to classify adults into two income brackets. A general description of the data is underlined available here. Do not download the data from this source, there are continuous and unknown values that have been removed in the dataset/stub file available to you on Mimir.

Each line of the file looks like this:
workclass, education, marital-status, occupation, relationship, race, sex, native-country, class-label

Using the provided stub file, implement a decision tree classifier using the **ID3 algorithm** (see the slides or the book). You should implement the train() function before the classify() function. These two functions are required as written in order to test your code. You are free to implement a decision tree however you wish, as long as you write the algorithm from scratch. Also, be sure not to edit the function signatures, or you will break the tests. You can write any extra functions that you deem necessary. **Do not use premade packages such as scikit-learn.**

We have provided a convert() function that turns each attribute vector into an integer vector. Please base the decisions in your model on the values of these integers, not the categorical values themselves. Both train() and classify() will be passed integer attribute vectors during tests on Mimir.

Notes:

- The value "Never-worked" of attribute *workclass* is removed from the training set when it is groomed of unknown values ("?"). If someone has never worked, then we don't know what their *occupation* is.

Assigned on 4/14/2017 11:04:00 AM | Due on 4/28/2017 11:59:00 PM   500 Submissions Remaining

Looks like you don't have any submissions yet
Click submit to submit your code!