Bronco ID: 013001924
Last Name: Nalbandian
First Name: Andre

Assignment 1

Friday, September 10, 2021   7:17 PM

1. ETP refers to the Experience Task and Performance of an ML system. The E is the experience the system gains as it continues to learn and loop through. The T is the task as a whole and what the system is trying to accomplish. The P is the performance or the accuracy of the ML system being correct.

    Using a Spam Vs Ham example. The T is correctly identifying what emails are spam vs ham. The E is the process of learning with email inputs what can be classified as spam and ham. The P is the accuracy of identifying emails as spam or ham.
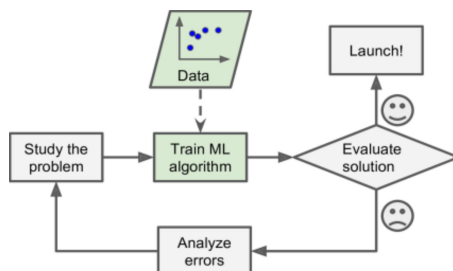
    For this diagram, the whole system is trying to achieve the Task of X. The Experience it gains is from the data being input to looping all the way around to train the ML algorithm. The performance is described when the solution is evaluated and the algorithm is launched.

2. The three main phases are Preprocessing, Machine Learning/Data Mining, and Post Postprocessing. They probably did this to simplify to simplify the explanation of the process. The first 4 steps and about 70-80% of the ML flow happens in preprocessing. This is where data is compiled and transformed to allow the machine to learn from good data and features. Garbage data produces garbage results regardless of the ML algorithm used. The ML phase is where the algorithm is trained and the data is input. Postprocessing is the process of visualizing the data output by the ML algorithm to allow for easy understanding.

3. Graphs
    a. Data Distribution - The training data set differs greatly from the test set.
    b. Outlier - 1 point in data is vastly different from the rest or the general trend of the data.
    c. Missing Values - When lots of data is left unknown.
    d. Noise - Irrelevant data that interrupt the pattern recognition of an ML algorithm
    e. Sparsity - When lots of data values are left to 0

4. The Contact Lens Data
    a. If a person needs to be recommended contact lenses.
    b. A feature is a characteristic or a property. Using this data an example of a feature is Age
    c. A feature value is the different values a certain feature can have. Example - Young/ Reduced/ Myope etc.
    d. Dimensionality is the number of features in the data set. The dimensionality for this data is 4.
    e. An instance is a collection of attributes that describe an object. In this graph each row is a new instance.
    f. A class is the values of the predictions made. This graph is a binary classifier so the 2 classes are Yes and No

5.
    a. Supervised - Learn a mapping from inputs x to outputs y given a labeled set of input-output pair $D = \{x_i, y_i\}_i^N = 1$. All data labels shown.

b. Unsupervised - Goal is to find "interesting patterns in the data" with only inputs $D_2 = \{(x_i)\}_i^N = 1$. No data labels shown.

c. Semi-Supervised - Learn a mapping from inputs x to outputs y given partially labeled data. Some data labels shown, usually more unlabeled than labeled.

6. A Binary classifier has 2 classes and 1 of the classes has to be assigned to each instance.
   A multi-class classifier has more than 2 classes but 1 of the classes has to be assigned to each instance.
   A multi-label classifier has more than 2 classes but each instance can be assigned to 0 or more classes.

7.

   a. H = [Prepresbyopic, Myope, No, Normal]
      = [?, Myope, ?, Normal]
      = [?, Myope, ?, Normal]
   H = [?, Myope, ?, ?]

   b. https://github.com/NootCode/4210-Assignment1.git

   c.

   d. https://github.com/NootCode/4210-Assignment1.git