015755428
Kim
Gary

1. A program that learns with experience, E, does so through data. It learns from the dataset to accomplish some goal; this would be the task, T. The accuracy of the machine's performance of this task is the performance measure P.
   In relation to the diagram, the experience is feeding your machine learning algorithm the dataset so that it can learn from it. The task is for the machine to find a solution to the problem. The performance measure would be the amount of errors the machine makes in its solution of the problem.

2. They probably did that because there is a clear distinction between the 3 phases. The 3 phases are preprocessing, running the machine learning algorithm, and postprocessing. In other words, the 3 phases are:
   - Getting your data and preparing it for your machine learning algorithm.
   - Running your machine learning algorithm.
   - Preparing the output of the ML algorithm so that humans can make meaningful insights.

   To build knowledge base, we need to gather data and prepare it for the ML algorithm. Then the ML algorithm should be run on the dataset to detect any and all patterns. Then the output of the algorithm needs to be formatted or displayed in ways that a human can make interpretations and learn something new.

3. a) data distribution. The image shows that the training set is centered around a mean of 1 while the test set is centered around a mean of 2. Because of this difference in distribution, the algorithm may perform poorly on the test set.

   b) outliers. This is a situation where most of the data are grouped nicely to form a pattern, but some data set values are far from the rest of the data.

   c) missing values. This can result from not being able to collect all the attributes of a specific instance or when certain attributes are not applicable. For example, the size of the garage does not apply to houses with no garage.

   d) noise. This is a result of data that does not fit a clean pattern.

   e) sparcity. Given features c1-c5 with a range of [0, 5], there are 6^5 possible combinations of feature values. However, because there is a limited collection of data, not all combinations may appear.

4. a) This is a classification task and they are trying to predict whether someone should get lenses given their age and the condition of their eyes.

   b) An feature is an attribute of an object. In this data, the attributes are age, spectacle prescription, astigmatism, tear production rate.

   c) A feature value is the value that a feature can take. For example, in this data, a feature value for astigmatism is yes or no.

   d) Dimensionality is the size of the feature vector. This data has a dimensionality of 4.

   e) An instance is a collection of attributes that describes a specific object in the data. For example, the first instance listed in the data is {young, myope, no, reduced}.

   f) A class is a value that you are trying to predict, this would be the dependent variable. In this data, it would be “Recommended Lenses”.

5. a) Since all data is labeled, supervised learning system should be used for clustering.
   b) Since no data is labeled, unsupervised learning system should be used.
   c) Since a few instances are labeled while the rest are unlabeled, semisupervised system should be used.

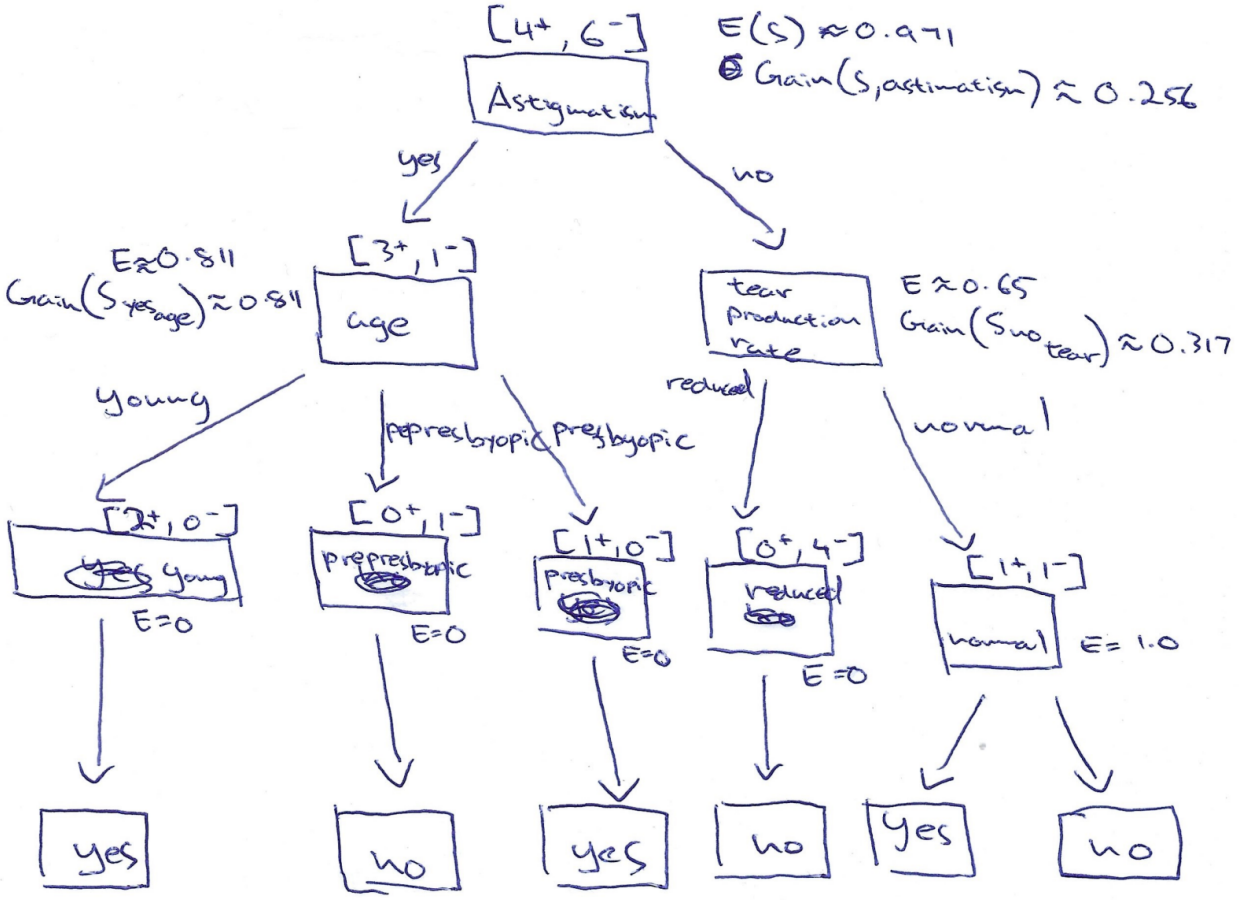
6. In the binary classifier, there are two possible labels and each object must be assigned one of these possible labels.

   In the multiclass classifier, there are a set of possible labels and each object must be assigned one of these labels. Note, the labels are mutually exclusive. For example, a fruit can be an orange or a banana, but not both.

   In the multilabel classifier, there are a set of possible labels and each object can be assigned any number of these labels. This is possible because the labels are not mutually exclusive. For example, A document can be about religion and politics or it can be about none of the labels in the class set.

7. a) h0 = (?, Myope, ?, ?)
   b) https://github.com/garyjsk271/machine-learning-class/tree/main/assignment%201

   c)



   d) https://github.com/garyjsk271/machine-learning-class/tree/main/assignment%201