CS4210 Assignment #1

Bronco ID: 015099318
Last Name: Greer
First Name: Cleighton

1. Experience "E" relates to the data and training components of the diagram. As the dataset provided to the machine changes, and the model as a result of training changes, then we say the model has recieved a new experience E. Performance "P" then correlates to the evaluate solution component as measurements are made to determine if changes in the data/outcome of training experience "E" produced a more accurate solution to task "T". "T" therefore is related to process of evaluating the solution, where "P" is the result of that process.

After obtaining result "P" that does not meet expectations, we analyze errors that occurred from "E" and find ways to improve "P" by better understanding the problem or the Task "T" to produce a better experience "E" to the machine.

2. KDD: Three phases

1. Preprocessing: Acquire data and transform and clean it.
   Without preprocessing, the data will be difficult to learn from. Preprocessing ensures that the data available with be in the best quality possible to discover patterns or other correlations by standardizing, minimizing noise and outliers, only selecting relevant features, handling missing values and sparsity in data, and minimizing dimensionality.

2. Machine Learn: Processing of data to create new models (or improve them)
   This is the step where data is processed for hidden meaning. Data alone is just data, in order to learn from it, various machine learning algorithms can be used to create models that reveal or unlock patterns within.

3. Post-processing: Analysis and evaluation of data.
   Now that the data has been factored into a model, we must interpret what patterns are discovered, why are they there? And what can be learned from these discoveries.

3.

a.) Image demonstrates the challenge of distribution.
The test data and the training data have mismatched distributions which will likely create a model that is biased towards the training data and innacurrately expect the same from the test data.

b.) Image demonstrate the challenge of outliers
Data instances that lie outside regular measurements are outliers, and may throw off ML algorithms ability to recognize common trends. Such as the curved line presented by the non-outlier data.

c.) Image demonstrates challenge of missing values
ML requires a lot of instances, but often not all features are accounted for in every instance. This hurts the quality of data when potential correlations are missing

d.) Image demonstrates challenge of noise
Similar to outliers, noise may throw off an algorithms ability to recognize clear trends. But noise does not present as extreme measurements, instead as fuzzy or meaningless instances that pollute the sample of otherwise concise and clear data points.

e.) Image demonstrates challenge of sparsity
Many instances are unique, with little repetition between them. While all features may be accounted for, a combination of feature values only occurs a single or very few times - this is the case for a majority of instances. This is a problem because it becomes difficult to relate instances to others and discern patterns, when each one is unique in its feature values.

(4)

a.) Determine whether a patient is compatible with contact lenses

b.) A feature is a characteristic or attribute measured across all instances.
For example each collumn in the data dictates a feature, such as age.
All instances had an age determined

c.) A feature value is a measurement or outcome that appears in a given feature.
For example, the feature value "no" appears in the Astigmatism Column for instances 1, 2, 3, 4, 7, and 9 of the data set. Therefore "no" is a feature value of Astigmatism

d.) Dimensionality simply describes the amount of features considered in the whole data set.
For example, since there are 4 columns with features in the data (not including the Recommended Lenses Column which is actually the Output/Class label)
then the dimensionality of the data is 4

e.) An Instance is a singular data point that comprises one measurement of each feature (and includes the outcome).
For example, each row (not including the header) represents an instance.
Row #1 measures a datapoint with a young age, Myope spectacle prescription, no astigmatism, and a reduced tear production rate. The output for this instance was that lenses were not recommended

f. A class is the distinguishing outcome that instances are separated into.
For Example, this data presents a binary classification of the instances into either a yes: recommended lenses class, or a no: recommended lenses class
The class of Instance 1 is "no".

5.)

a.) Supervised Machine Learning should be used in this scenario

- all data is clearly labled with distinct classes

b.) Unsupervised Machine Learning should be used in this scenario to describe how the data is clustured

- The data is all labled the same so we can assume there are no lables however there is still some distinct separation between groupings of data

c.) Semi-Supervized Machine Learning is best used in this scenario

- most of the data is indistinguishable because they share a generic-label, so similarly to part b.), we can say it is unlabled data. However there are a select few labled data points that seem to fit a trend as in part a.) so we can use a Semi-Supervised Technique to take advantage of partially labled data and infer lables on the rest of the data

6.)

Binary Classifier:

Out of $K=2$ classes, assign $C=1$ number of classes to a given instance
(Select 1 class out of 2)

Multi-label Classifier:

Out of $K>2$ classes, assign $C \geq 0$ number of classes to a given instance
(Select 0 or more classes out of 3 or more)

Multiclass Classifier:

Out of $K>2$ classes, assign $C=1$ number of classes to a given instance
(Select only 1 class out of 3 or more classes)

$$Gain(S,A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

7.)

a.) $\boxed{\text{Root: Astigmatism}}$

$$\underline{Entropy(S)} = -\frac{4}{10}\log_2\frac{4}{10} - \frac{6}{10}\log_2\frac{6}{10} = \underline{0.9709}$$

$$Gain(S, Age) = 0.9709 - \sum_{v \in values(Age)} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$= 0.9709 - \left(\frac{4}{10} Entropy(young) + \frac{3}{10} Entropy(presbyopic) + \frac{3}{10} Entropy(prepresbyopic)\right)$$

$$\frac{4}{10} Entropy(young) = \frac{4}{10} \cdot 1 = 0.4$$

$$\frac{3}{10} Entropy(presbyopic) = \frac{3}{10}\left(\frac{1}{3}\log\frac{1}{3} - \frac{2}{3}\log\frac{2}{3}\right) = \frac{3}{10}(0.9183) = 0.2755$$

$$\frac{3}{10} Entropy(prepresbyopic)\ \frac{3}{10}\left(-\frac{1}{3}\log\frac{1}{3} - \frac{2}{3}\log\frac{2}{3}\right) = 0.2755$$

$$= 0.9709 - .9509 = \boxed{0.0199}$$

$$Gain(S, Spec\ Prescription) = 0.9709 - \sum_{v \in values(SP)} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$= 0.9709 - \left(8/10\ Entropy(Myope) + \frac{2}{10} Entropy(Hypermetrope)\right)$$

$$8/10\ Entropy(Myope) = 8/10 \cdot 1$$

$$2/10\ Entropy(Hypermetrope) = 2/10 \cdot 0$$

$$= 0.9709 - 0.8 = \boxed{0.1709}$$

$$Gain(S, Astigmatism) = 0.9709 - \sum_{v \in values(Astigmatism)} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$= 0.9709 - \left(6/10\ Entropy(No) + 4/10\ Entropy(yes)\right)$$

$$6/10\ Entropy(No) = 6/10\left(-\frac{1}{6}\log_2\frac{1}{6} - \frac{5}{6}\log_2\frac{5}{6}\right) = \frac{6}{10}(0.6500) = 0.3900$$

$$4/10\ Entropy(yes) = 4/10\left(-3/4\log_2 3/4 - \frac{1}{4}\log_2\frac{1}{4}\right) = 4/10(0.8113) = 0.3245$$

$$= 0.9709 - (0.39 + 0.3245) = \boxed{0.2564}$$

$$Gain(S, Tear\ Prod\ Rate) = 0.9709 - \sum_{v \in values(Tear\ Rate)} \frac{|S_v|}{|S|} Entropy(S_v)$$

$$= 0.9709 - \left(6/10\ Entropy(Reduced) + 4/10\ Entropy(Normal)\right)$$

$$6/10\ Entropy(Reduced) = 6/10\left(-1/6\log_2 1/6 - 5/6\log_2\frac{5}{6}\right) = \frac{6}{10}(0.6500) = 0.3900$$

$$4/10\ Entropy(Normal) = 4/10\left(-3/4\log_2 3/4 - \frac{1}{4}\log\frac{1}{4}\right) = 0.3245$$

$$= 0.9709 - (0.39 + 0.3245) = \boxed{0.2564}$$

$\boxed{L1\ (No): Age}$

$$Entropy(S(No)) = 0.65$$

$$Gain(S_{(no)}, Age) = 0.65 - \sum_{v \in Values(Age)} \frac{|S_{(no)v}|}{|S_{(no)}|} Entropy(S_{(no)v})$$

$$= 0.65 - \left(\tfrac{2}{6}(Entropy(young)) + \tfrac{2}{6}(Entropy(presbyopic)) + \tfrac{2}{6}(Entropy(prepresbyopic))\right)$$

- $\frac{2}{6} Entropy(young) = \frac{2}{6}(0) = 0$
- $\frac{2}{6} Entropy(presbyopic) = \frac{2}{6} \cdot 0$
- $\frac{2}{6} Entropy(prepresbyople) = \frac{2}{6} \cdot 1$

$$= 0.65 - (0 + 0 + \tfrac{2}{6}) = \boxed{0.3167}$$

$$Gain(S_{(no)}, Spec\ Prescription) = 0.65 - \sum_{v \in Values(Spec.P.)} \frac{|S_{(no)v}|}{|S_{(no)}|} Entropy(S_{(no)v})$$

$$= 0.65 - \left(\tfrac{4}{6} Entropy(Myope) + \tfrac{2}{6} Entropy(Hypermetrope)\right)$$

- $\frac{4}{6} Entropy(Myope) = \frac{4}{6}\left(-\frac{1}{4}\log_2 \frac{1}{4} - \frac{3}{4}\log_2 \frac{3}{4}\right) = \frac{2}{3}(0.8113) = 0.5408$
- $\frac{2}{6} Entropy(Hypermetrope) = \frac{2}{6}(0) = 0$

$$= 0.65 - (0.5408) = \boxed{0.1091}$$

$$Gain(S_{(no)}, Tear\ production\ Rate) = 0.65 - \sum_{v \in Values(TPD)} \frac{|S_{(no)v}|}{|S_{(no)}|} Entropy(S_{(no)v})$$

$$= 0.65 - \left(\tfrac{4}{6} Entropy(Reduced) + \tfrac{2}{6} Entropy(Normal)\right)$$

- $\frac{4}{6} Entropy(Reduced) = \frac{4}{6}(0)$
- $\frac{2}{6} Entropy(Normal) = \frac{2}{6}(1)$

$$= 0.65 - (\tfrac{2}{6}) = \boxed{0.3167}$$

$\boxed{L1\ (Yes): Age}$

$$Entropy(S(yes)) = 0.8113$$

$$Gain(S_{(yes)}, Spec\ Prescription) = 0.8113 - \sum_{v \in Values(SP)} \frac{|S_{(yes)v}|}{|S_{(yes)}|} Entropy(S_{(yes)v})$$

$$= 0.8113 - \left(\tfrac{4}{4} Entropy(Myope) + \tfrac{0}{4} Entropy(Hypermetrope)\right)$$

- $\frac{4}{4} Entropy(Myope) = 1 \cdot \left(-\frac{3}{4} Entropy(\frac{3}{4}) - \frac{1}{4}(Entropy(\frac{1}{4}))\right) = 0.8113$

$$= \boxed{0}$$ (no information gain)

$$Gain(S_{(yes)}, Age) = 0.8113 - \sum_{v \in Values(Age)} \frac{|S_{(yes)v}|}{|S_{(yes)}|} Entropy(S_{(yes)v})$$

$$= 0.8113 - \left(\tfrac{2}{4} Entropy(young) + \tfrac{1}{4} Entropy(presbyopic) + \tfrac{1}{4} Entropy(prepresbyopic)\right)$$

- $\frac{2}{4} Entropy(young) = \frac{2}{4} \cdot 0$

$-\frac{1}{4}$ Entropy (presbyopic) $= \frac{1}{4} \cdot 0$
$-\frac{1}{4}$ Entropy (prepresbyopic) $= \frac{1}{4} \cdot 0$
$= 0.8113 - 0 = \boxed{0.8113}$ (max information gain)

$$\text{Gain}(S_{(yes)}, \text{Tear Prod Rate}) = 0.8113 - \sum_{v \in \text{Values(TPD)}} \frac{|S_{(yes)v}|}{|S_{(yes)}|} \text{Entropy}(S_{(yes)v})$$

$= 0.8113 - \left(\frac{2}{4} \text{Entropy(Normal)} + \frac{2}{4} \text{Entropy(Reduced)}\right)$

$- \frac{2}{4}$ Entropy(Normal) $= \frac{2}{4}(0)$
$- \frac{2}{4}$ Entropy(Reduced $= \frac{2}{4}(1)$
$= 0.8113 - \left(0 + \frac{2}{4}\right) = \boxed{0.3113}$

Astigmatism: "No"

$\boxed{\text{L2(Young): Class label = "no"}}$
Entropy(S_{young}) = 0 ∴ leaf node of majority class

$\boxed{\text{L2(presbyopic): Class label = "no"}}$
Entropy($S_{presbyopic}$) = 0 ∴ leaf node of majority class

$\boxed{\text{L2(prepresbyopic): Class label = "No"}}$
Entropy($S_{prepresbyopic}$) = 1 ∴ leaf node of previous majority class (Astigmatism: No)

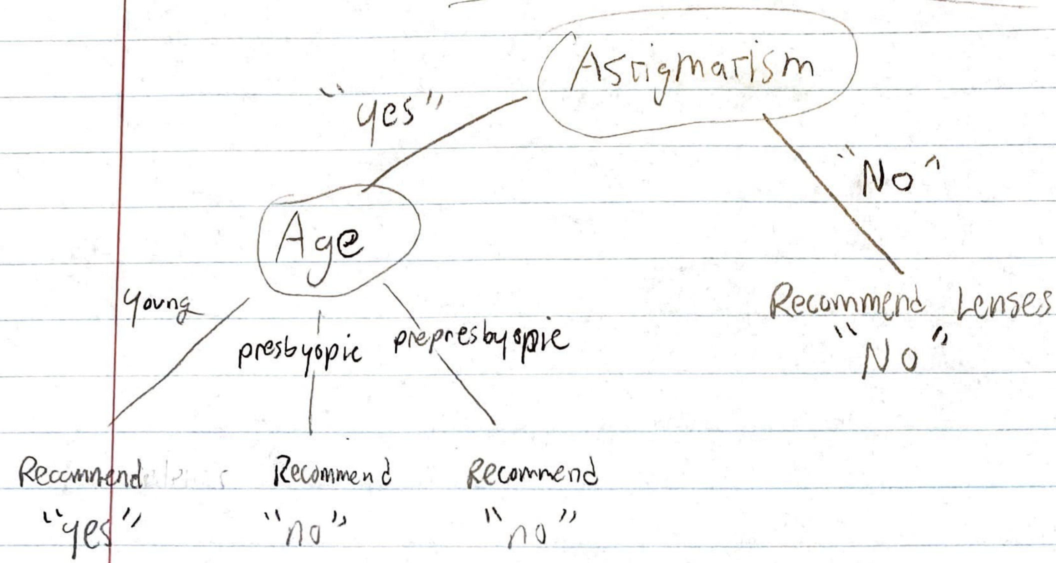
Astigmatism "yes"

$\boxed{\text{L2(Young): class label = "yes"}}$
Entropy($S_{(young)}$) = 0 ∴ leaf node of majority class

$\boxed{\text{L2(prebyopic): Class label = "no"}}$
Entropy($S_{(prebyopic)}$) = 0 ∴ leaf node of majority class

$\boxed{\text{L2(prepresbyopic): class label: "no"}}$
Entropy($S_{(prepresbyopic)}$) = 0 ∴ leaf node of majority class

ID3 Decision Tree



C. The trees differ because the algorithm used (In part b.) can only create a binary split at each node. ID3 is not limited to a binary split.