Bronco ID: | 0 | 1 | 5 | 2 | 6 | 2 | 6 | 2 | 4 |

Last Name: Nguyen

First Name: Loc

1. Experience E is the training data being fed to the ML algorithm. Task T is going to solve the problem being studied and performance P is the result of the evaluation solution step. If performance is good, algorithm can launch. Otherwise, bad performance means that the solution needs to be analyzed & studied for errors.

2. There are 3 phases – preprocessing, machine learning, post-processing.

   Preprocessing deals with handling missing values, bias in the data, the curse of dimensionality through aggregation, feature selection, or dimensionality reduction for sparse data. This phase is about preparing data for ML to prevent garbage in, garbage out.

   Machine learning is when the algorithm learn from the data through super, semi, or unsupervised or reinforcement learning. Depending on the type of tasks (classification, regression prediction, etc.), different types of learning method is selected or combined to provide the best solution.

   Postprocessing is the analysis phase and data is interpreted & visualized so trends and patterns can help draw conclusions to a problem that the scientist is researching.

3a. The training data sample isn’t representative of the test data sample. Distribution problem.

3b. Outlier in data can cause skewed data

3c. Missing values in the data sample makes the data incomplete and hard to use

3d. There is noise in the data as seen with random deviation throughout a particular time slot leads to loss of info about the trend in the data

3e. Sparsity in the data

4a. Trying to figure out if a person needs lenses

4b. A feature is an attribute that is taken into data collection to draw a conclusion. Attributes for this dataset can be the presence of astigmatism or tear production rate

4c. A feature value is the measured value or outcome of the attribute. Value can be binary or multiple. For example, the feature value for astigmatism is “yes” or “no”

4d. Dimensionality is the number of features that the dataset contain. In this case, the dataset has four dimensions: age, spectacle prescription, astigmatism, and tear production rate.

4e. An instance is an object of the class. In this case, the instance in the dataset is each individual person included in the data collection

4f. A class is a blueprint or a collection of objects based on common features. In this case, the class is people who wear contact lens

5a. Supervised learning should be used because the sample has clear data labels

5b. Unsupervised learning should be used because the sample doesn’t have data labels and the machine will try to draw boundaries on its own

5c. Semi-supervised learning should be used because the algorithm is dealing with a little bit of labeled data and a lot of unlabeled data, trying to map from inputs x to outputs y

6.

Multiclass classifier has C = 1 meaning that an object can only be assigned to one class (or one label) and K isn't mentioned but the number of classes could be 1 or greater.

Multilabel classifier has C greater or equal to 0 meaning that an object may be assigned to more than 1 class / label or doesn't have a label or class at all. The number of classes K must be greater than 2 for object to possess more than 1 label. Also, it's the superset of multiclass label.

Binary classifier has C = 1 meaning than an object can only assigned to one class (binary meaning one class or the other). The number of classes K = 2 so that "bi" meaning only two classes are permitted.

7a.

Note; log = log base 2

First split:

Entropy(S) = -0.4log(0.4) – 0.6log(0.6) = 0.971

Gain (S, Age) = 0.971 – 0.4*Entropy(S young) – 0.3*Entropy(S prebyopic) – 0.3*Entropy(S preprebyopic) = 0.971 – 0.4*1 – 0.3*0.918 – 0.3*0.918 = 0.02

Gain(S, Spectacle Prescription) = 0.971 – 0.8*Entropy(S myope) – 0.2*Entropy(S hypermetrope) = 0.971 – 0.8(1) – 0.2(0) = 0.171

Gain(S, Astigmatism) = 0.971 – 0.4*Entropy(S yes) – 0.6*Entropy(S no) = 0.971 – 0.4*0.811 – 0.6*0.65 = 0.257

Gain(S, Tear Production Rate) = 0.971 – 0.4*Entropy(S normal) – 0.6*Entropy(S reduced) = 0.971 – 0.4*0.811 – 0.6*0.65 = 0.257

Since information gain for astigmatism and tear production rate is same, we can pick randomly between the two attributes. I pick astigmatism as the root.

Second split for has astigmatism:

Gain (S has astigmatism, S age) = 0.811 – 0.5*(0) – 0.25*0 – 0.25*0 = 0.811

Gain (S has astigmatism, S spectacle prescription) = 0.811 – 1*E(has astigmatism, myope) – 0*E(has astigmatism, hypermetrope) = 0.811 – 1*0.811 = 0

Gain(S has astigmatism, S Tear Production Rate) = 0.811 – 0.5*E(has astigmatism, normal) – 0.5*E(has astigmatism, reduced) = 0.811 – 0.5*0 – 0.5*1 = 0.311

Since information gain is maximized if age is selected, I pick age as the next attribute under has astigmatism.

Third split for no astigmatism:

Gain (S no astigmatism, S spectacle prescription) = 0.65 – (4/6)*E(no astigmatism, myope) – (2/6)*E(no astigmatism, hypermetrope) = 0.65 – 2/3*0.811 – 1/3*0 = 0.109

Gain (S no astigmatism, S tear production rate) = 0.65 – (4/6)*E(no astigmatism, reduced) – (2/6)*E(no astigmatism, normal) = 0.65 – 2/3*0 – 1/3*1 = 0.317

Since information gain is maximized if tear production rate is selected, I pick tear production rate as the next attribute under no astigmatism.

Fourth split for spectacle prescription:

Gain(S has astigmatism, young, spectacle prescription) = 0 – 1*0 = 0

Gain(S has astigmatism, prebyopic, spectacle prescription) = 0

Gain(S has astigmatism, preprebyopic, spectacle prescription) = 0

Gain(S no astigmatism, normal TPR, spectacle prescription) = 1 – (2/2)*E(S no astigmatism, normal TPR, myope) = 1 – 1*1 = 0

Gain(S no astigmatism, reduced TPR, spectacle prescription) = 0

Since information gain is none with spectacle prescription added, the attribute is discard.

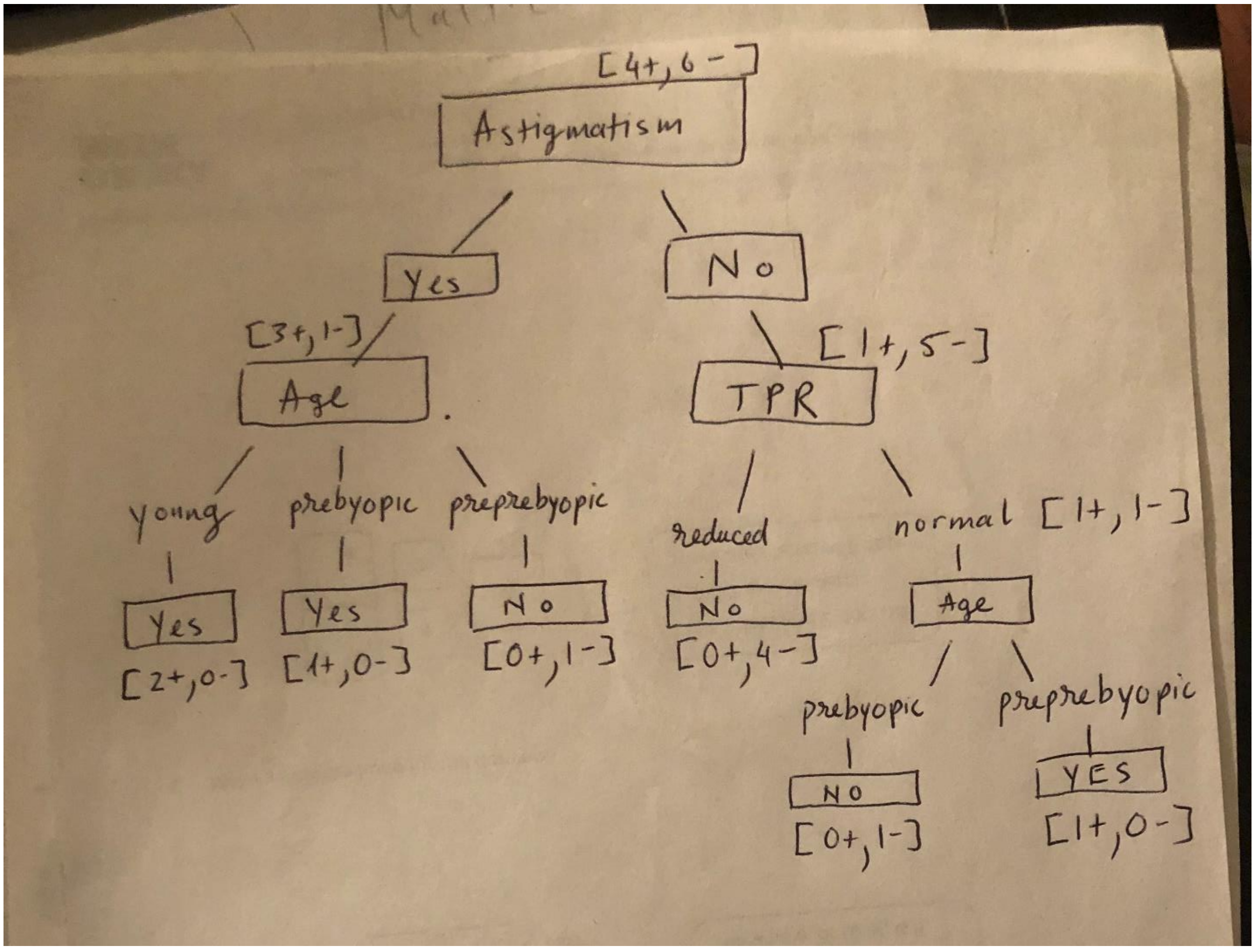
Fifth split for no astigmatism, normal TPR:

Gain(S no astigmatism, normal TPR, spectacle prescription) = 1 – 2/2*E(S no astigmatism, normal TPR, myope) = 1 – 1*1 = 0

Gain(S no astigmatism, normal TPR, age) = 1 – 1/2*E(S no astigmatism, normal TPR, presbyopic) – 1/2 * E(S no astigmatism, normal TPR, preprebyopic) = 1 – 0.5*0 – 0.5*0 = 1

Since IG is maximized with age, age is selected under no astigmatism, normal TPR

The ID3 decision we drawn is as below:



7b. https://github.com/Skyhorizon2021/CS_4210/blob/main/decision_tree.py

7c. The decision tree is different because of the randomization factor when picking attribute when information gain for either attribute is the same. Instead of Astigmatism as the root, we can also have tear production rate as the root.