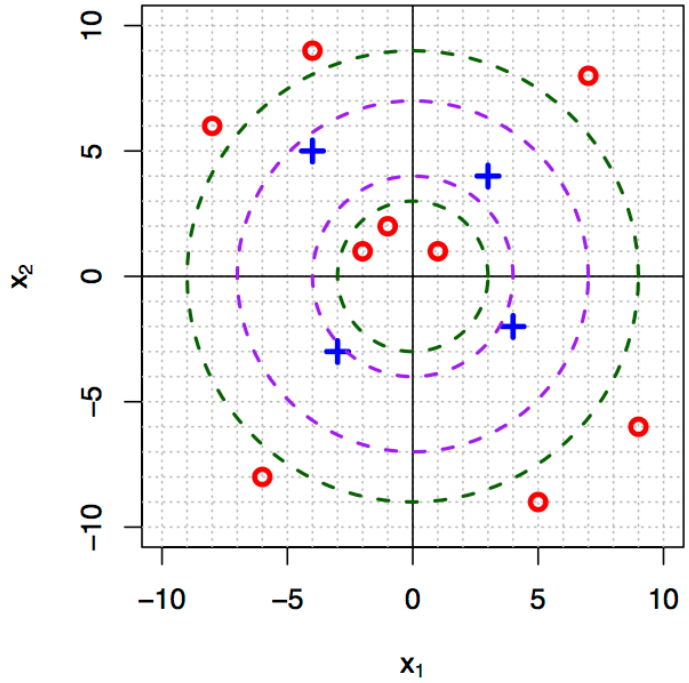
CS6140: Machine Learning – Homework 1
Name: Vishal Rajpal

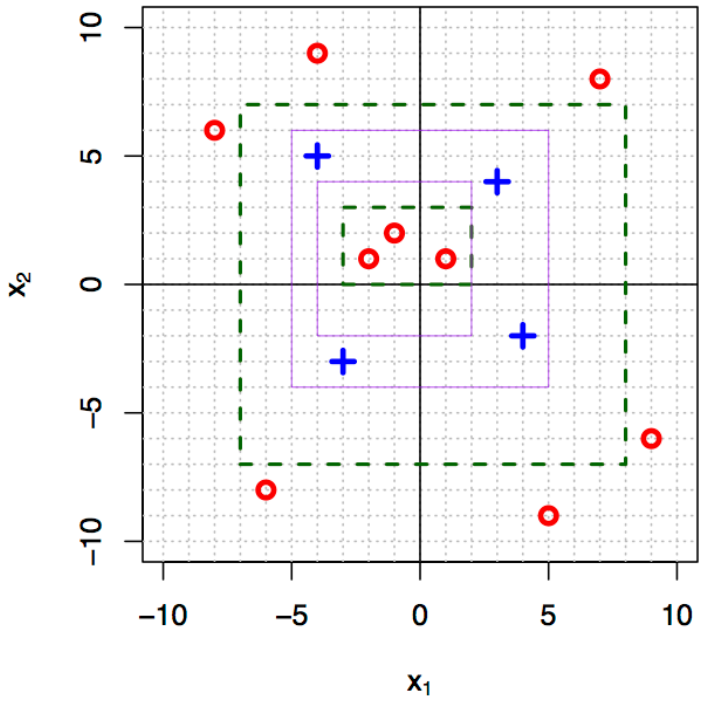
Problem 1: Version Spaces

a) The below figure illustrates the S and G boundaries. The concentric circles with purple color represent S and the concentric circles with green color represent G.



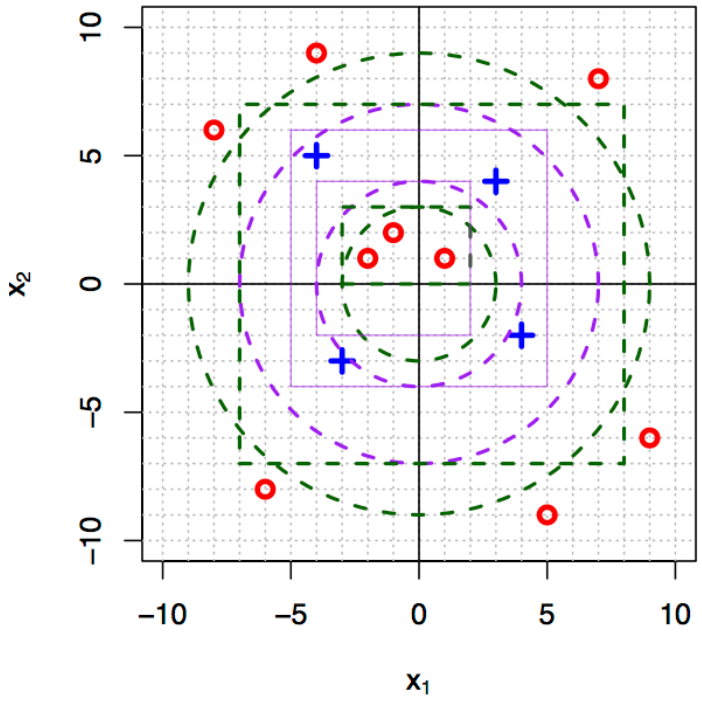
Based on the figure the version space will be **[(3,7), (3,8), (3,9), (4,7), (4,8), (4,9)].** Hence size of the version space is **6**.

b) The below figure illustrates the S and G boundaries. The rectangles with purple color represent S and the rectangles with green color represent G.



$\mathcal{H}_2$ can be defined as a set of 4 points/coordinates where two coordinates belong to the outer rectangle and 2 belong to the innser rectangle. The two points represent the bottom left point and top right point of the rectangle. Based on the figure above $\mathcal{H}_2$ hypotheses can be of form $a_1 < x < b_1$where (a_1, a_2) is the bottom left coordinate of the outer rectangle and (b_1, b_2) is the top right coordinate of the inner rectangle. This representation is similar to that of circle. The version space based on this figure will be the number of rectangles at various coordinates inside the S and G boundaries.

c) The below figure represents both the version spaces



Based on the above figure lets suppose we have a point (5,6). This point will shrink the circle version space as the S boundary will have its outer circle move more outward if the label is positive and if the label is negative the outer circle of G boundary will have to move inward. Hence irrespective of the label the version space will shrink. Same inference stands for rectangle version space too.

To shrink only one version space the point should be outside one version space but inside the other. For example point (-8,0) will shrink the circle version space only.

d) $\mathcal{H}_1$has less hypotheses space than $\mathcal{H}_2$ and hence $\mathcal{H}_1$will generalize better than $\mathcal{H}_2$. As the hypotheses space is small there will be less overfitting of the data.

Problem 2: Decision Tree

All the below calculations will be based on the training data.

Let A be the event that the instance belongs to label 1: $\boldsymbol{\mathcal{P}(A) = \frac{3}{18} = \frac{1}{6}}$

Let B be the event that the instance belongs to label 2: $\boldsymbol{\mathcal{P}(B) = \frac{3}{18} = \frac{1}{6}}$

Let C be the event that the instance belongs to label 3: $\boldsymbol{\mathcal{P}(C) = \frac{12}{18} = \frac{2}{3}}$

Entropy = $\mathcal{H}(S_l) = \sum_{i=1}^{|y|} -p(i)\log_2 p(i)$

Therefore,

$$\mathcal{H}\left(\frac{1}{6},\frac{1}{6},\frac{2}{3}\right) = -\frac{1}{6}\ln\left(\frac{1}{6}\right) - \frac{1}{6}\ln\left(\frac{1}{6}\right) - \frac{2}{3}\ln\left(\frac{2}{3}\right)$$

$$= -0.166 * \ln(0.166) - 0.166 * \ln(0.166) - 0.666 * \ln(0.666)$$

$$= 0.298 + 0.298 + 0.270$$

$$\mathcal{H}\left(\tfrac{1}{6},\tfrac{1}{6},\tfrac{2}{3}\right) = 0.866$$

Gain (Age)

Let A be the event such that the instance belongs to label 1

Let B be the event such that the instance belongs to label 2

Let C be the event such that the instance belongs to label 3

Let H be the event that the Age = 1

Therefore,

$$\mathcal{P}(A|H) = \frac{1}{6}$$

$$\mathcal{P}(B|H) = \frac{2}{6} = \frac{1}{3}$$

$$\mathcal{P}(C|H) = \frac{3}{6} = \frac{1}{2}$$

$$\mathcal{H}\left(\frac{1}{6},\frac{1}{3},\frac{1}{2}\right) = -\frac{1}{6}\ln\left(\frac{1}{6}\right) - \frac{1}{3}\ln\left(\frac{1}{3}\right) - \frac{1}{2}\ln\left(\frac{1}{2}\right)$$

$$= -0.166 * \ln(0.166) - 0.333 * \ln(0.333) - 0.5 * \ln(0.5)$$

$$= 0.298 + 0.366 + 0.346$$

$$\mathcal{H}\left(\frac{1}{6},\frac{1}{3},\frac{1}{2}\right) = 1.01$$

Let H be the event that the Age = 2

Therefore,

$\mathcal{P}(A|H) = \frac{1}{6}$

$\mathcal{P}(B|H) = \frac{1}{6}$

$\mathcal{P}(C|H) = \frac{4}{6} = \frac{2}{3}$

$$\mathcal{H}\left(\frac{1}{6},\frac{1}{6},\frac{2}{3}\right) = -\frac{1}{6}\ln\left(\frac{1}{6}\right) - \frac{1}{6}\ln\left(\frac{1}{6}\right) - \frac{2}{3}\ln\left(\frac{2}{3}\right)$$

$$= -0.166 * \ln(0.166) - 0.166 * \ln(0.166) - 0.666 * \ln(0.666)$$

$$= 0.298 + 0.298 + 0.270$$

$$\mathcal{H}\left(\frac{1}{6},\frac{1}{6},\frac{2}{3}\right) = 0.866$$

Let H be the event that the Age = 3

Therefore,

$\mathcal{P}(A|H) = \frac{1}{6}$

$\mathcal{P}(B|H) = 0$

$\mathcal{P}(C|H) = \frac{5}{6}$

$$\mathcal{H}\left(\frac{1}{6},0,\frac{5}{6}\right) = -\frac{1}{6}\ln\left(\frac{1}{6}\right) - \frac{5}{6}\ln\left(\frac{5}{6}\right)$$

$$= -0.166 * \ln(0.166) - 0.833 * \ln(0.833)$$

$$= 0.298 + 0.152$$

$$\mathcal{H}\left(\frac{1}{6},0,\frac{5}{6}\right) = 0.45$$

E[Age] = $\frac{1}{3} * 1.01 + \frac{1}{3} * 0.866 + \frac{1}{3} * 0.45$

$= 0.336 + 0.288 + 0.15$

$= 0.774$

$Gain(Age) = H(S) - E(Age)$

$= 0.866 - 0.774$

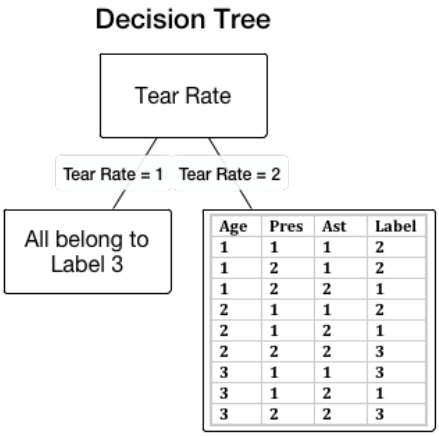
$= 0.092$

Similarly

$Gain(Prescription) = 0.048$
$Gain(Astigmatic) = 0.229$
$\boldsymbol{Gain(Tear\ Rate) = 0.316}$

Therefore, we gain maximum by splitting on Tear Rate at the first level. When we split on Tear Rate all the elements with TearRate = 1 have label 3.

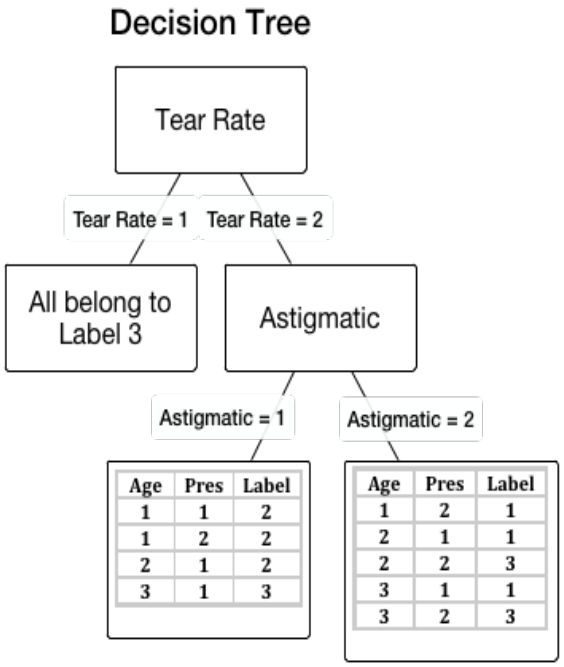
Hence the tree by splitting till level 1 looks like below:



Age	Pres	Ast	Label
1	1	1	2
1	2	1	2
1	2	2	1
2	1	1	2
2	1	2	1
2	2	2	3
3	1	1	3
3	1	2	1
3	2	2	3

$Gain(Age|TR = 2) = 0.308$
$Gain(Prescription|TR = 2) = 0.049$
$\boldsymbol{Gain(Astigmatic|TR = 2) = 0.477}$

Hence we split on Astigmatic. The resulting tree looks like as below:



Age	Pres	Label
1	1	2
1	2	2
2	1	2
3	1	3

Age	Pres	Label
1	2	1
2	1	1
2	2	3
3	1	1
3	2	3

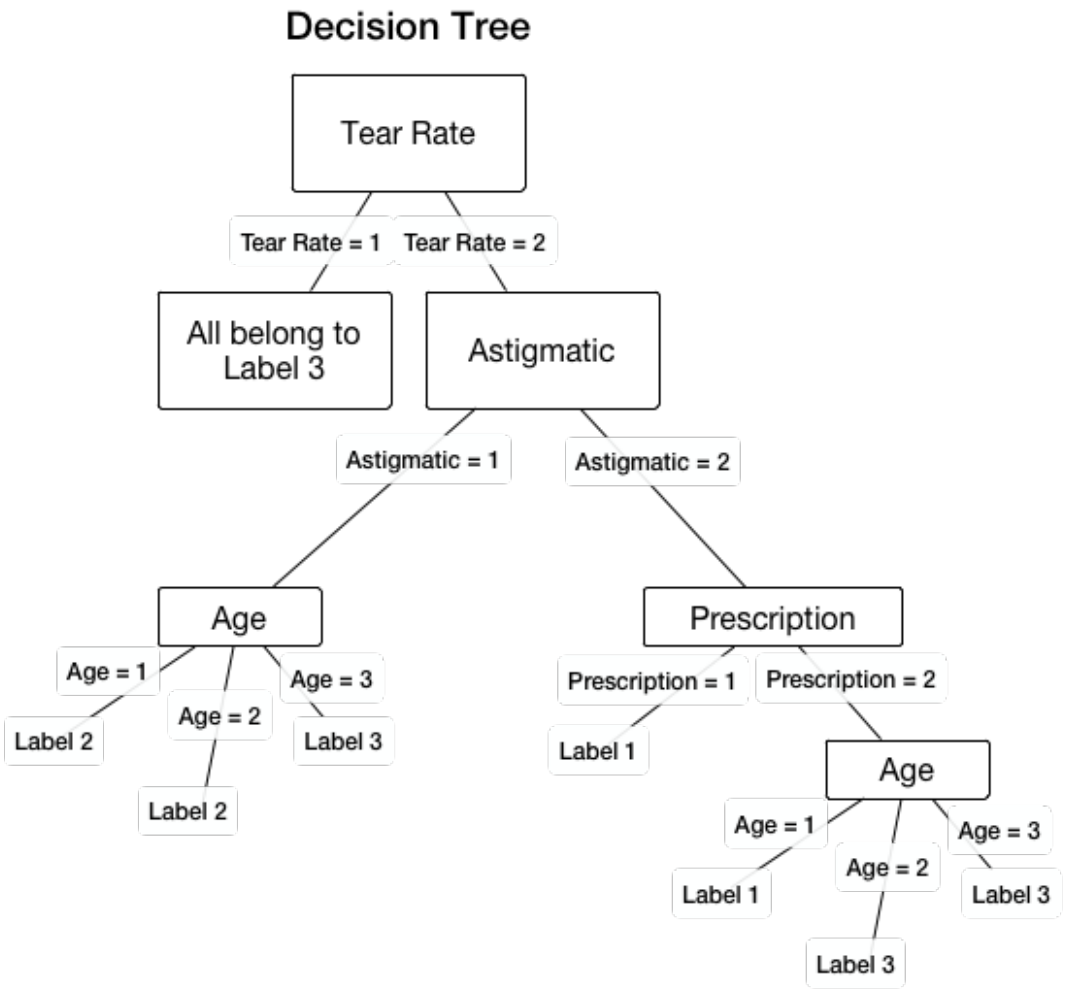
$\boldsymbol{Gain(Age|TR = 2|Astigmatic = 1) = 0.561}$
$Gain(Prescription|TR = 2|Astigmetic = 1) = 0.084$

Therefore, we split on Age on left side of Astigmatic split.

$Gain(Age|TR = 2|Astigmatic = 2) = 0.12$
$\boldsymbol{Gain(Prescription|TR = 2|Astigmatic = 2) = 0.291}$

Therefore, we split on Prescription on right side of Astigmatic split.

The final decision tree is as below:



Id	Age	Prescription	Astigmatic	Tear Rate	Label	Prediction Label
3	1	1	2	1	**3**	**3**
4	1	1	2	2	**1**	**1**
11	2	1	2	1	**3**	**3**
14	2	2	1	2	**2**	**2**
17	3	1	1	1	**3**	**3**
22	3	2	1	2	**2**	**3**

Hence Accuracy $= \frac{5}{6}$

Problem 3: Nearest Neighbor (https://github.com/vishalrajpal/MachineLearning**)**

a) Approach: Preprocessing

Step 1: Read the training and testing file and represent in our own class structure.
Step 2: **Merge the testing data with the training data** to calculate mean and other needed values on the entire set.
Step 3: **Normalize** the attribute values using **z-scale technique** with **ignoring the instances that have a missing value.**
Step 4: **Impute** the missing values by **label conditioned Mean/Median** imputation technique.
Step 5: Produce the **processed** files with the normalized and imputed data.

b) Approach: K-nearest neighbor

Step 1: Read the training and testing file and represent it in our own class structure.
Step 2: Compute the **L2 distance** for each testing instance with all training instances.
Step 3: **Sort them in ascending order of the distance for each testing instance.**
Step 4: Read the **first k** training instances from the sorted data.
Step 5: Assign the **majority label in these k instances** to the **testing instance**.
Step 6: After the process is complete for every instance output the testing data to Standard Output.

c) Accuracy Report

File / k	$\boldsymbol{k = 3}$	$\boldsymbol{k = 10}$
lenses.testing	$\frac{4}{6} = 0.67$	$\frac{3}{6} = 0.5$
crx.testing.processed	$\frac{117}{138} = 0.84$	$\frac{118}{138} = 0.85$