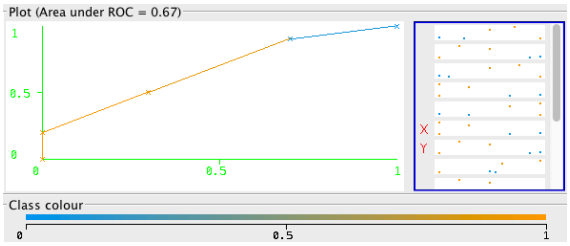
CS760 – Assignment 1

Data set: The aim of the project is to detect gender after extracting a set of features from Repose Frontal View and Repose Lateral View of facial images of both Males and Females between age 25-30 years. Some of the features extracted based on study done on human faces of men and women are:

1. **Frontal Bossing** Ratio which is Cranial Width divided by Cranial Length - For male the ratio will be higher than female in general.
2. **Eyebrows Curvature** which is the distance between the lowest point of the lower eyebrow and highest point of lower eyebrow - Males have straight eyebrows as compared to females.
3. **Eyebrow Width** - Males have comparatively broader eyebrows as compared to females in general.
4. **Eyes Narrowness** - Males have narrow and flatter eyes unlike females who have more of oval shape.
5. **Eyelid Length** - Females have longer eyelid than males without taking cosmetics into picture.
6. **Length of nose** - Males have an average length of 5.8 cm and females have 5.1 cm.
7. **Width of Nose** - Males have a wider nose as compared to females in general.
8. **Ratio of Upper lip by lower lip** - Males have comparatively thin upper lip.
9. **Chin angle** - Men have lesser jaw to chin angle than females who have sharper chin.
10. **Neck length to Neck Width** - Men have comparatively shorter neck as compared to females.

J48 TREE

```
J48 pruned tree
------------------

Width of Upper Lip by Lower Lip <= 0.64: Male (8.0)
Width of Upper Lip by Lower Lip > 0.64
|   Nose Protruding Length cm <= 2.78: Female (10.0)
|   Nose Protruding Length cm > 2.78: Male (2.0)

Number of Leaves  :     3

Size of the tree :      5
```



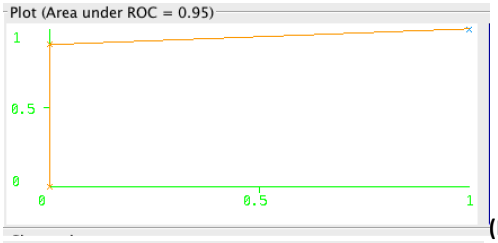

```
=== Detailed Accuracy By Class ===

                 TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                   0.5       0.3       0.625      0.5       0.556      0.67     Male
                   0.7       0.5       0.583      0.7       0.636      0.67     Female
Weighted Avg.      0.6       0.4       0.604      0.6       0.596      0.67

=== Confusion Matrix ===

 a b   <-- classified as
 5 5 | a = Male
 3 7 | b = Female
```

The tree created initially (as seen in ADTree) was bigger with more number of leaves and had higher accuracy with 14 correct and 6 incorrect classifications but after pruning, the tree is much smaller and some of the ambiguities in data like where men face features are more like women or vice versa got misclassified. The pruning came at the cost of accuracy. As seen in confusion matrix 5 males and 3 females were misclassified by J48. AUC in ROC is also lower.

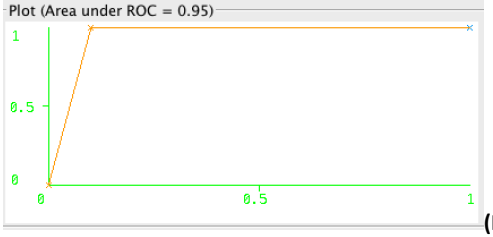
IB1

```
=== Detailed Accuracy By Class ===

                 TP Rate   FP Rate   Precision   Recall  F-Measure   ROC Area  Class
                   0.9       0          1         0.9       0.947      0.95     Male
                   1         0.1        0.909     1         0.952      0.95     Female
Weighted Avg.      0.95      0.05       0.955     0.95      0.95       0.95

=== Confusion Matrix ===

  a  b   <-- classified as
  9  1 |  a = Male
  0 10 |  b = Female
```

The accuracy is higher and true positive rate is 0.9. Only 1 male is being misclassified as female as also seen in confusion matrix. The ROC AUC is high. Through results it looks like every instance has a neighbor which will classify



(Male)



(Female)

But if we increase the number of instances or take in to consideration images from different ethnic groups, then there will be a need of more features. While analyzing J48 tree I realized that the data was being classified only based on nose length. I realized the images I had for training set were biased towards that feature and there can be more instances, which I am not considering, that can give me different results When I changed the training data then I started getting lesser accuracy in tree classifier and realized there will be need of more features to classify properly.