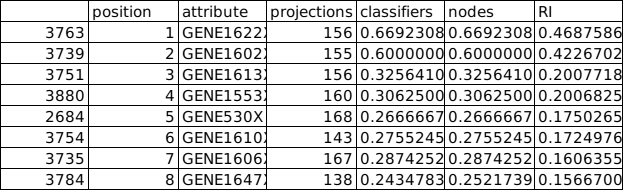
1. What was the background of the data?
2. What kind of the data was used? How was the data obtained? Evaluate quality and quantity of the dataset.
3. What were the most important features?

Cutoff value = 0.13224232



1. What was the aim of the classification model?
2. What kind of models and algorithms did you use? Explain why.

Used MCFS to find the relevant features to explain the data.

Did pre-discretisation then and given as input for.

Used both johnson and genetic reducer with rosetta.

1. Which parameters did you choose? Explain why.

Initial data:

108 objects with 11002 attributes.

Projectionsize = 105

Projections = 10477

Splits = 5

cutoffPermutations = 20

For mcfs.

mcfs\_result <- mcfs(decision~., data, projections=auto,

projectionSize=auto, splits=5, splitSetSize=0.66,

cutoffPermutations = 20, threadsNumber = 8)

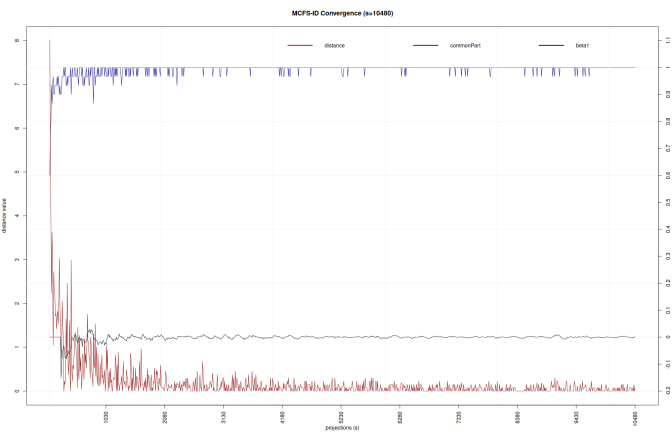
For rosetta.

rosetta\_out3 <- rosetta(input\_rosetta,discrete = TRUE, roc = TRUE, clroc = "Age28plus")

rosetta\_genetic\_out3 <- rosetta(input\_rosetta,discrete = TRUE,reducer="Genetic", roc = TRUE, clroc = "Age28plus")

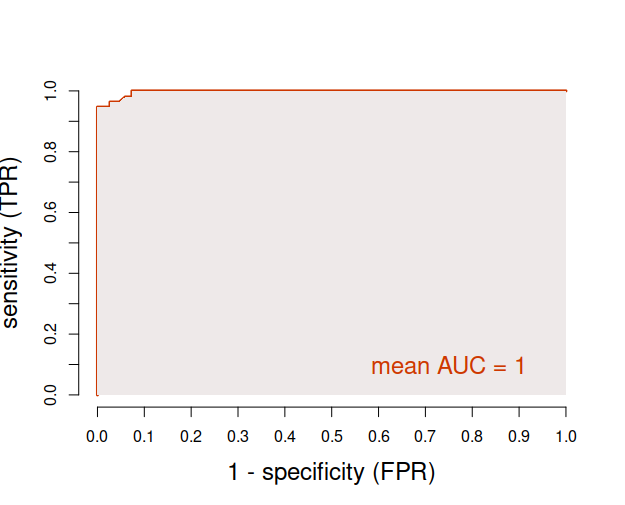
1. How did you evaluate your models?

MCFS using distance plot

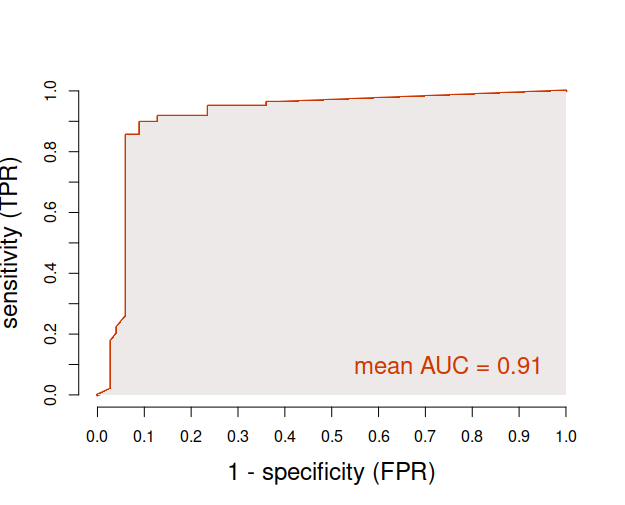


For rosetta it is by using visunet for plotting and validating the connections and nodes. And using the accuracy

The decision class with more number is used to produce ROC curve by runnning rosetta again.

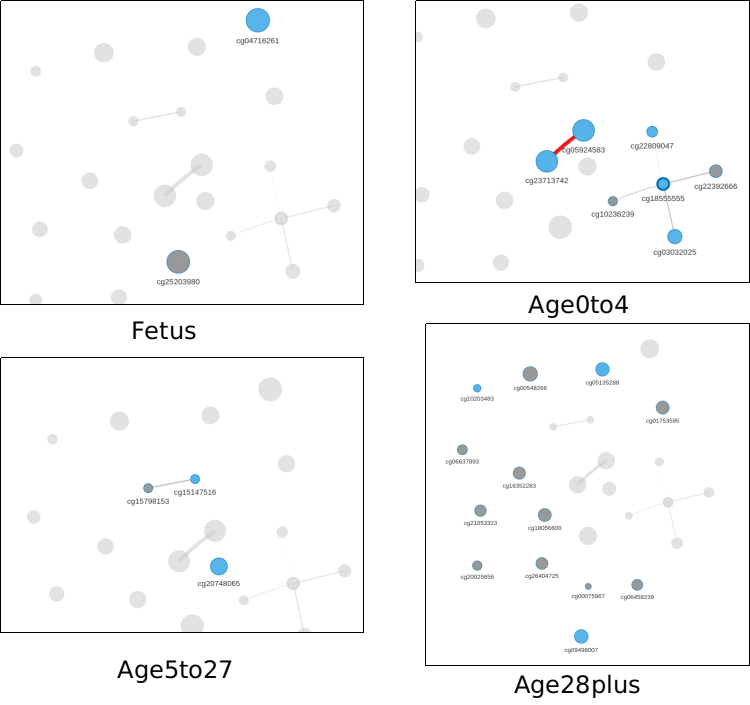


Genetic roc Age28plus



Johnson roc Age28plus

1. Give an interpretation of the VisuNet graphs. Show the strongest connections, hubs or separations etc.



Strongest connection is seen in Age0to4 for cg05924583 and cg23713742

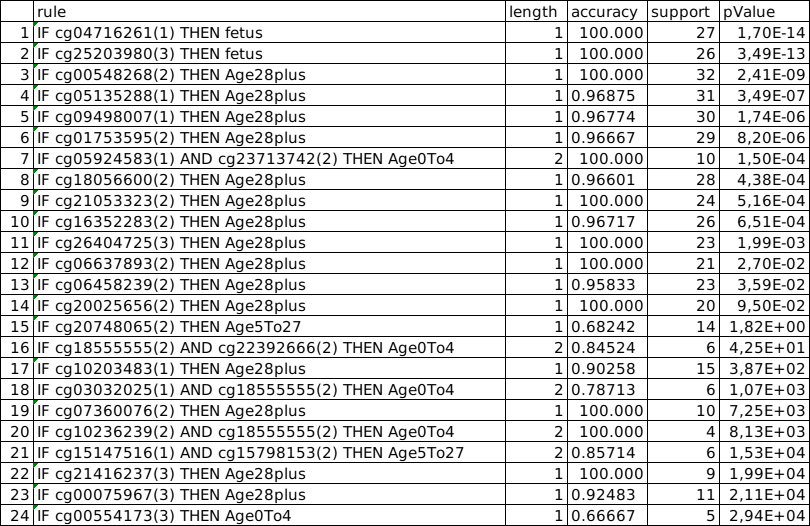
cg05924583 and cg23713742 have 10 connection with 1 edge with each other which is the stongest with the johnson algorithm.

With reducer being genetic, the network was dense that I was expensive to even calculate. So the rules are filtered with 0.01 pValue and then sorted according to their support and coverage. Then from each decision class 50 top rules are plotted using visunet.

1. Report the top rules for each class and provide a brief interpretation of them.

(p=0.05)

All the significant rule from johnson



There are 18936 significant rules in genetic.

1. Report the features that discern between different classes (from the top 5 rules).
2. Investigate the top rules for common features. Compare the results with the VisuNet network.
3. What results did you get? What was the performance? Which methods worked best?

Genetic reducer worked best in rosetta with more number of connections, but accuracy is reduced.

Rosetta johnson mean = 0.777778

Rosetta genetic mean = 0,748485

1. Were the models themselves informative? Give an example of what you can learn from the models.
2. What was difficult or could be improved in the project methodology?