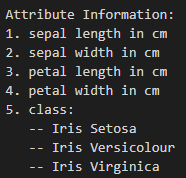
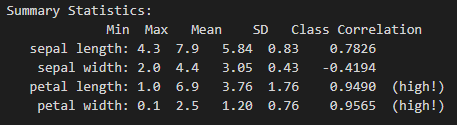
**Programming assignment 2 Report Dimitrije Prosevski**

**Problem Statement and Description**

The objective of the programming assignment is to compare two data analysis methods using Iris dataset. Iris data set is a contains three different species: Setosa, Versicolor, Virginica. I will be using decision tree and neural network for this assignment. They will learn upon the data and classify them based off of their attributes. All the coding is done in RStudio (R language).



**Brief Description of the Two Methods Used**

**Decision tree:** The idea of decision tree is to start with all the variables, then separate them by some attribute that best separates the group, and then splitting the groups even more and the process repeats until the partitions. In other words, decision tree is like a flowchart where nodes represent the comparisons done on the attributes, branches represent the outcomes of the comparisons, and leaf nodes represent a class label. Classification rules are obtained when going root to leaf node.

**Neural network:** Neural network is an imitation of human brain in a sense, because it works by having a network of neurons (nodes) that are connected to each other and intercommunicate with stimulating signals. Each neuron is a node and for every node there is an attribute variable that has some weight applied to it. Neurons are organized in layers where every neuron from one level connects to every neuron in the next level.

**Experimental Results (code in blue)**

**Decision tree:**

data(iris)

names(iris)

#install.packages("rpart")

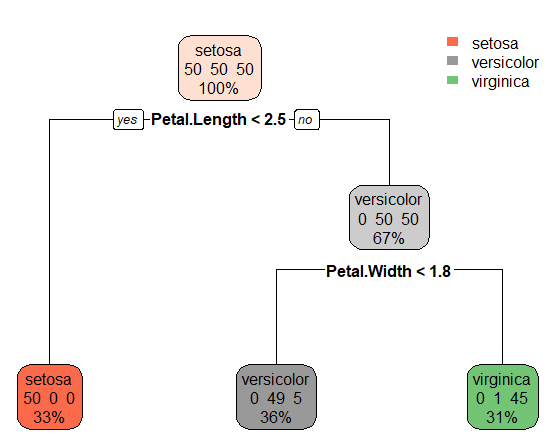
#install.packages("rpart.plot")

library(rpart)

library(rpart.plot)

myTree <- rpart(formula = Species ~ ., iris, method = "class")

rpart.plot(myTree, type=2, extra=101)



**Neural network:**

library(neuralnet)

set.seed(101)

size.sample <- 50

iristrain <- iris[sample(1:nrow(iris), size.sample),] # get a training sample from iris

nnet\_iristrain <- iristrain

nnet\_iristrain <- cbind(nnet\_iristrain, iristrain$Species == 'setosa')

nnet\_iristrain <- cbind(nnet\_iristrain, iristrain$Species == 'versicolor')

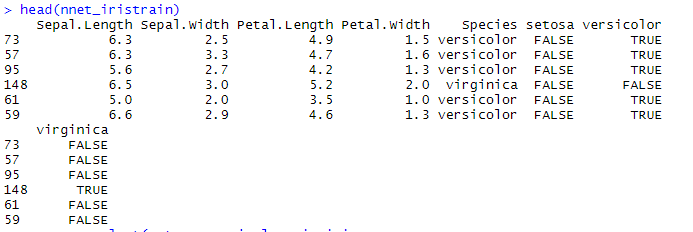
nnet\_iristrain <- cbind(nnet\_iristrain, iristrain$Species == 'virginica')

names(nnet\_iristrain)[6] <- 'setosa'

names(nnet\_iristrain)[7] <- 'versicolor'

names(nnet\_iristrain)[8] <- 'virginica'

head(nnet\_iristrain)

****

**Discussion of Results**

**Decision tree:**

From the decision tree we can see that first split happened on comparison “petal length < 2.5” if yes, its most likely Setosa, if no its either versicolor or virginica. Now on the right side another split happens depending on if “petal width < 1.8”. If petal width smaller than 1.8 then its most likely to be versicolor, if greater than 1.8 its virginica.

In this scenario, versicolor predicted correct class in 49 cases but missed 5 cases where versicolor got classified as virginica.

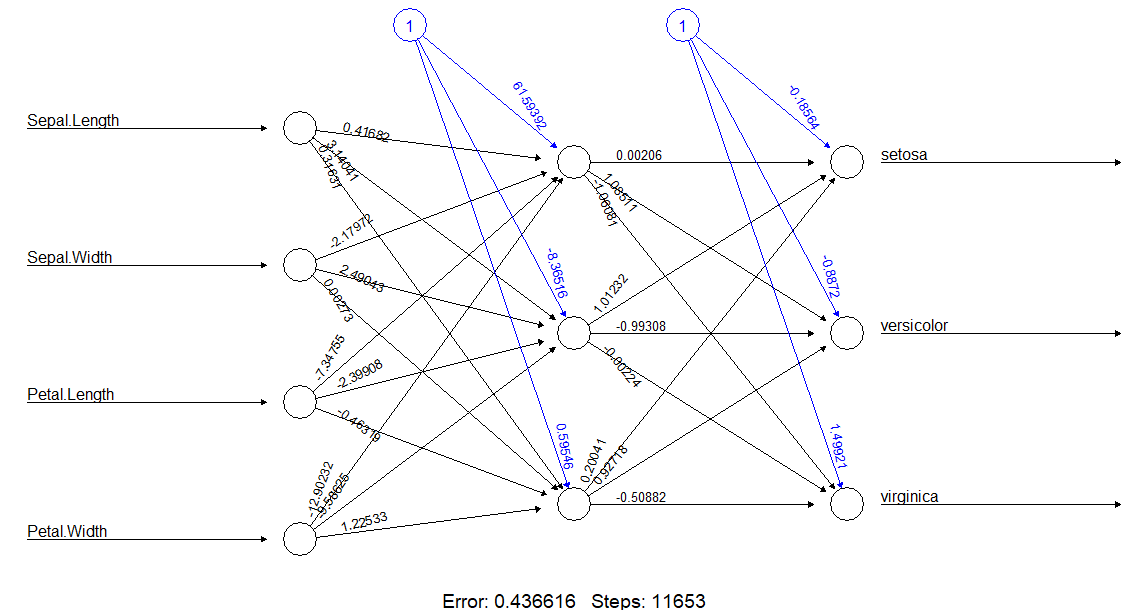
So, out of 150 samples there have been 6 errors (versicolor 5 error, virginica 1 errors)

If petal length less than 2.5 then its most likely Setosa, if petal length greater than 2.5 and petal width smaller than 1.8 its versicolor, if petal length greater than 2.5 and petal width greater than 1.8 its virginica.

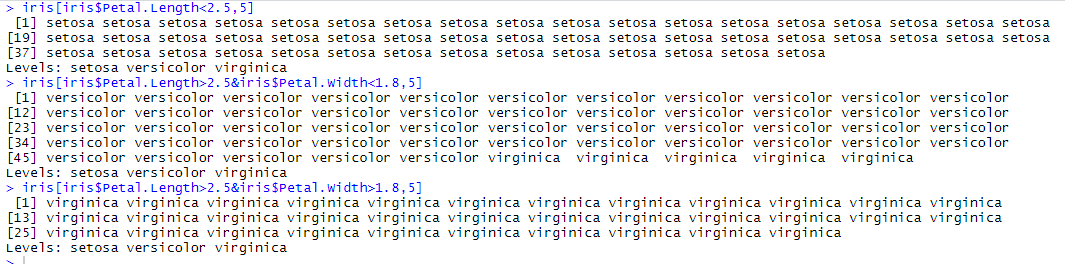
**Neural network:**

nn <- neuralnet(setosa+versicolor+virginica ~ Sepal.Length+Sepal.Width+Petal.Length+Petal.Width, data=nnet\_iristrain, hidden=c(3))

plot(nn)

**Conclusion:** Error rate for decision tree is 0.04 and for neural network is 0.44. Decision tree is less accurate than neural network but more interpretable. In this case, using decision tree was faster to execute and draw a plot since there are way less steps in comparison to neural network which has 11653 steps. Neural networks are better for binary data than decision trees, however they cannot handle categorical values.

**Proof:**

****

**References:**

<https://www.youtube.com/watch?v=JFJIQ0_2ijg>

<https://davetang.org/muse/2013/03/12/building-a-classification-tree-in-r/?fbclid=IwAR1HhmtAqlipivtJKd6g1f95TYj6YUldYV3SJVO-KK3ow6xosF3vzFc5h2A>

<https://rpubs.com/ksevong/392295?fbclid=IwAR1hym18J5SQ47MMs3ZgqRJs8CIXxzD75XMywhNsAID4z7OZ0JKBVvH2bb4>