STAT 652 - Assignment 2

Dhruv Patel, 301471961

Question 1

1. Fit a default Random Forest (RF) to only the three main variables in the data—Temp, Wind, and Solar.R—and not the two extra ones that we engineered. A RF should be able to detect interactions automatically if needed.

```r
# Helper Functions
get.MSPE = function(Y, Y.hat){
  return(mean((Y - Y.hat)^2))
}

# Create k CV folds for a AQset of size n
get.folds = function(n, K) {
  ### Get the appropriate number of fold labels
  n.fold = ceiling(n / K) # Number of observations per fold (rounded up)
  fold.ids.raw = rep(1:K, times = n.fold) # Generate extra labels
  fold.ids = fold.ids.raw[1:n] # Keep only the correct number of labels
  # Shuffle the fold labels
  folds.rand = fold.ids[sample.int(n)]
  return(folds.rand)
}
```
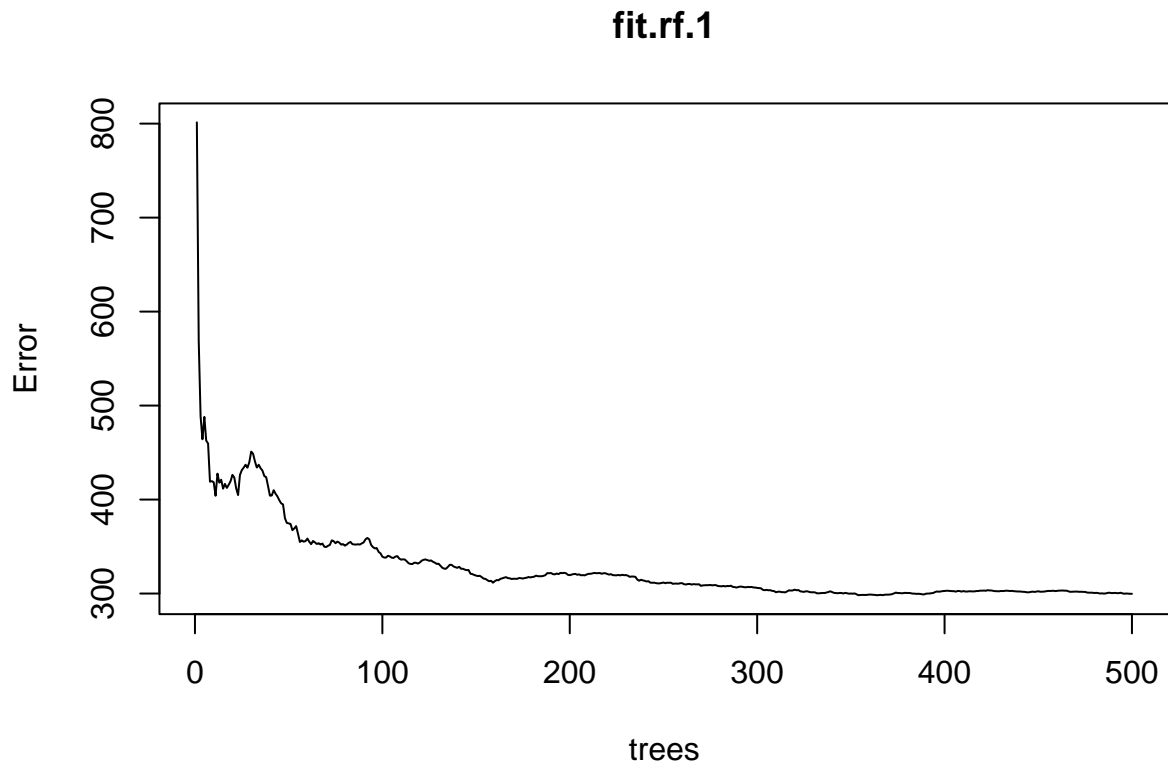
```r
AQ = airquality[1:4]

# Removing Null values
AQ = na.omit(AQ)

# importing libraries for Random Forest and setting setting seed
library(randomForest)
set.seed(301471961)

# Training the model
fit.rf.1 = randomForest(Ozone ~ ., data = AQ, importance = T)

# Plotting the model
plot(fit.rf.1)
```

fit.rf.1


```r
# Predict the model on Training set
OOB.pred.1 = predict(fit.rf.1)
```

1. (a) Report the OOB error.
   Answer: OBB error is 299.7023.

```r
# Get the Mean Square Prediction Error
OOB.MSPE.1 = get.MSPE(AQ$Ozone, OOB.pred.1)
OOB.MSPE.1
```

```
## [1] 299.7023
```

1.(b) Produce variable importance measures and comment on the relative importance of the variables.
How do they compare to what we have seen in earlier analyses of these data?
Answer: Based on the below importance measures Temp is the most important then Wind and at last
Solar.R. Comparing to the analysis done before we are getting similar variables as important feature i.e
(Temp > wind > Solar.R)

```r
# Model Summary
summary(fit.rf.1)
```

```
##              Length Class  Mode
## call            4    -none- call
## type            1    -none- character
```

```
## predicted       111    -none- numeric
## mse             500    -none- numeric
## rsq             500    -none- numeric
## oob.times       111    -none- numeric
## importance        6    -none- numeric
## importanceSD      3    -none- numeric
## localImportance   0    -none- NULL
## proximity         0    -none- NULL
## ntree             1    -none- numeric
## mtry              1    -none- numeric
## forest           11    -none- list
## coefs             0    -none- NULL
## y               111    -none- numeric
## test              0    -none- NULL
## inbag             0    -none- NULL
## terms             3    terms  call
```

```r
# Get Important variable from the model
importance(fit.rf.1)
```

```
##          %IncMSE IncNodePurity
## Solar.R 14.39503      24670.91
## Wind    23.16163      41382.91
## Temp    37.55516      46970.69
```

```r
# Plot Important variable
varImpPlot(fit.rf.1)
```

fit.rf.1