

STAT 652 - Assignment 2

Dhruv Patel, 301471961

Question 2

2. Repeat the exercise in question 1, adding the two engineered features into the data.

```
# 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)
}
```

2. (a) Report the OOB error. Does it improve much compared to the previous RF analysis without the variables?
Answer: Yes, OOB error (MSPE) has been reduced from 299.7023 to 280.3793 after including new featured variables.

```
AQ = airquality[1:4]

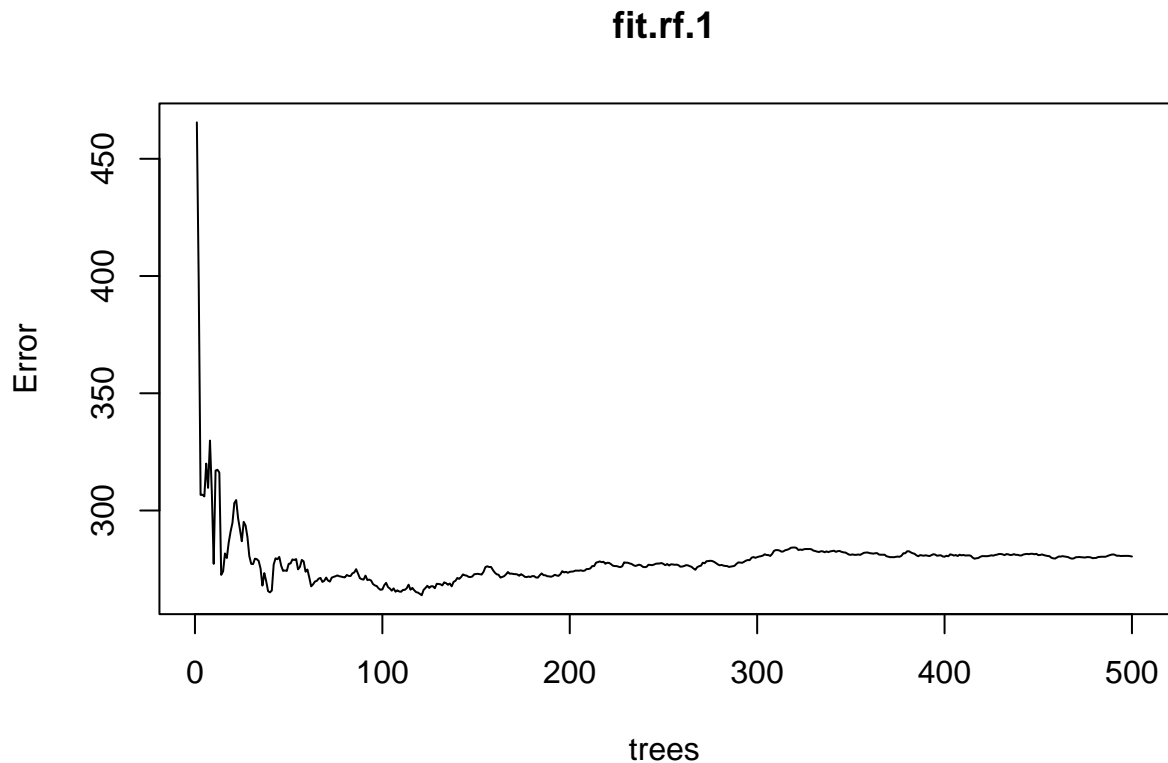
# Removing Null values
AQ = na.omit(AQ)

# Adding two new feature
AQ$TWcp = AQ$Temp*AQ$Wind
AQ$TWrat = AQ$Temp/AQ$Wind

# 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)
```



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

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

```
## [1] 280.3793
```

2. (b) Produce variable importance measures. Are the two engineered features particularly important?
 Answer: As per the output for importance measures, it seems like TWrat is significantly more important than TWcp. Also, adding them also reduce overall MSPE. So these engineered features are important.

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

```
##          %IncMSE IncNodePurity
## Solar.R 14.07299      14094.25
## Wind    14.65704      22294.79
## Temp    25.35940      29344.61
## TWcp     13.23687      19568.11
## TWrat    19.36180      29179.09
```

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

fit.rf.1

