Algorithm 3

Code

Hide

```
library(caret)
library(ranger)
library(tidyverse)
library(e1071)
library(readxl)
red <- read_excel("C:/Users/sarah/Desktop/info8000/winequalityred.xlsx",
    col_types = c("numeric", "numeric", "numeric",
        "numeric", "numeric", "numeric",
        "numeric", "numeric", "numeric",
        "numeric", "numeric", "numeric"))
library(caret)
library(ranger)
library(tidyverse)
library(e1071)
set.seed(65468)
inTraining<-createDataPartition(red$quality,p=.75,list = FALSE)
training<-red[inTraining,]
testing<-red[-inTraining,]
fitControl<-trainControl(## 10-fold CV
  method = "cv",
  number=10,
  ##repeated ten times
  repeats = 10)
```

```
`repeats` has no meaning for this resampling method.
```

Hide

```
rfFit1<- train(quality ~.,data=training,
               method="rf",
               trControl = fitControl)
rfFit1
```

```
Random Forest

1200 samples
  11 predictor

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 1080, 1079, 1080, 1081, 1082, 1080, ...
Resampling results across tuning parameters:

  mtry  RMSE       Rsquared   MAE
   2    0.5908197  0.4927346  0.4410006
   6    0.5928324  0.4795846  0.4372829
  11    0.5943479  0.4758715  0.4349772

RMSE was used to select the optimal
 model using the smallest value.
The final value used for the model was
 mtry = 2.
```

Hide

```
save(file='rffit.model',rfFit1)
```

Hide

```
rf2fit<- train(quality ~ alcohol + citricacid + volatileacidity + totalsulfurdioxide + sulphates + fixedacidity + density,
               data=training,
               method="rf",
               trControl=fitControl)
```

performs best when all variables are included!Still not great though…