STAT 602: Machine Learning

Homework 2

Spring 2014, Dr. Stephen Vardeman

Name:	Ian Mouzon
email:	imouzon@iastate.edu

Instructions:	Submit all code; show all methods
Assignment:	Problem 1 - 5
Due Date:	Monday May 4, 2015

I am using the following packages:

```
library(ggplot2)
library(lubridate)
library(xtable)
library(foreach)
library(rCharts)
library(magrittr)
library(tidyr)
library(dplyr)
library(reshape2)
library(gtools)
library(sqldf)
library(missForest)
```

and my working directory is set to `dmc2015/ian`.

0.1 Reading the Data

I am working from the current feature matrix:

```
featMat = readRDS("~/dmc2015/data/featureMatrix/featMat_v2.0.rds")
trn = featMat$train
cls = featMat$class

#Also reading the melted train and test sets
trn.m = read.csv("~/dmc2015/data/clean_data/melted_train_simple_name.csv")
cls.m = read.csv("~/dmc2015/data/clean_data/melted_test_simple_name.csv")

stack.trn = trn.m
stack.trn$dsn = "trn"

stack.cls = cls.m
stack.cls$dsn = "cls"

stack.m = rbind(stack.trn,stack.cls)

stack.m$dsn = factor(stack.m$dsn,levels=c('trn','cls'))
```

In case I need to reference the raw data, I will read that too:

```
raw.trn = read.csv("~/dmc2015/data/clean_data/train_simple_name.csv")
raw.cls = read.csv("~/dmc2015/data/clean_data/test_simple_name.csv")
```

0.2 Getting the results for all 2, 3, 4, and 5 way combinations of categories:

We can repeat what we did above for all 2-3-4-5 way combinations of these categories:
We now have two valuable tools: `catFreq` and `catCombFreq`