# ISYE6501_HW6

*June 27, 2018*

**Question 14.1**

The breast cancer data set breast-cancer-wisconsin.data.txt has missing values. 1. Use the mean/mode imputation method to impute values for the missing data. 2. Use regression to impute values for the missing data. 3. Use regression with perturbation to impute values for the missing data. 4. (Optional) Compare the results and quality of classification models (e.g., SVM, KNN) built using (1) the data sets from questions 1,2,3; (2) the data that remains after data points with missing values are removed; and (3) the data set when a binary variable is introduced to indicate missing values.

Parts 1-3:

```
bc <- read.table("14.1breast-cancer-wisconsin.dataSummer2018.txt", sep=",",
                 na.strings = "?", header=FALSE, stringsAsFactors = FALSE)

#Mean with Rounding Imputation
MeanImputeRounding=cbind(bc)
MeanImputeRounding[is.na(MeanImputeRounding)]=round(mean(bc$V7,na.rm=TRUE))

#Median Imputation
MedianImpute=cbind(bc)
MedianImpute[is.na(MedianImpute)]=median(bc$V7,na.rm=TRUE)

#Mode Imputation
getmode <- function(v) {
   uniqv <- unique(v)
   uniqv[which.max(tabulate(match(v, uniqv)))]
}
ModeImpute=cbind(bc)
ModeImpute[is.na(ModeImpute)]=getmode(bc$V7)


#Regression Imputation
RegressImpute=cbind(bc)
lm1 <- lm(V7 ~ ., data=RegressImpute)
for (i in 1:nrow(RegressImpute)){
  if (is.na(RegressImpute$V7[i])==TRUE){
    RegressImpute$V7[i]=predict(lm1, newdata=RegressImpute[i,], type="response")
  }}


#Regression with perturbation Imputation
RegressPerturbImpute=cbind(bc)
for (i in 1:nrow(RegressPerturbImpute)){
  if (is.na(RegressPerturbImpute$V7[i])==TRUE){
    RegressPerturbImpute$V7[i]=predict(lm1, newdata=RegressPerturbImpute[i,],
                                       type="response")+rnorm(1,0,1)
  }}
```

Imputation with MICE:

```
miceData<-complete(tempData,1)
```

(2) Missing values removed:

```
#NA's removed:
NAremoved <- bc[complete.cases(bc),]
```

(3) the data set when a binary variable is introduced to indicate missing values:

```
BinImpute=cbind(bc)
BinImpute$V12 = ifelse(is.na(BinImpute$V7), 1, 0)
```

### Question 15.1

Describe a situation or problem from your job, everyday life, current events, etc., for which optimization would be appropriate. What data would you need?

Answer: Optimization is all econ students do! :) Kidding. But anyway, let's say I wanted to construct a nice large cylindrical swimming pool in my backyard that held 1000 L of water. We could use optimization to determine the minimum amount of materials needed (not that this approach is really recommended, but you might get on HGTV or something!)

### References

https://www.tutorialspoint.com/r/r_mean_median_mode.htm