

## P8106\_Mtp

Ekta Chaudhary

26/03/2020

```
library(tidyverse)
library(readxl)
library(caret)
library(ModelMetrics)
library(glmnet)
library(gam)
library(mgcv)
library(splines)
library(pdp)
library(earth)
library(dplyr)
library(nanar)
library(bnstruct)
library(corrplot)
library(logisticPCA)
```

## Reading the Datasets

```
Diagnosis =  
read_csv('./data/Diagnosis.csv')  
Prognosis =  
    read_csv('./data/Prognosis.csv')  
Breast_Cancer =  
    read_csv(file = './data/Breast_Cancer.csv' ,col_names = c('id_number','Clump_thickness','Uniformity_o
```

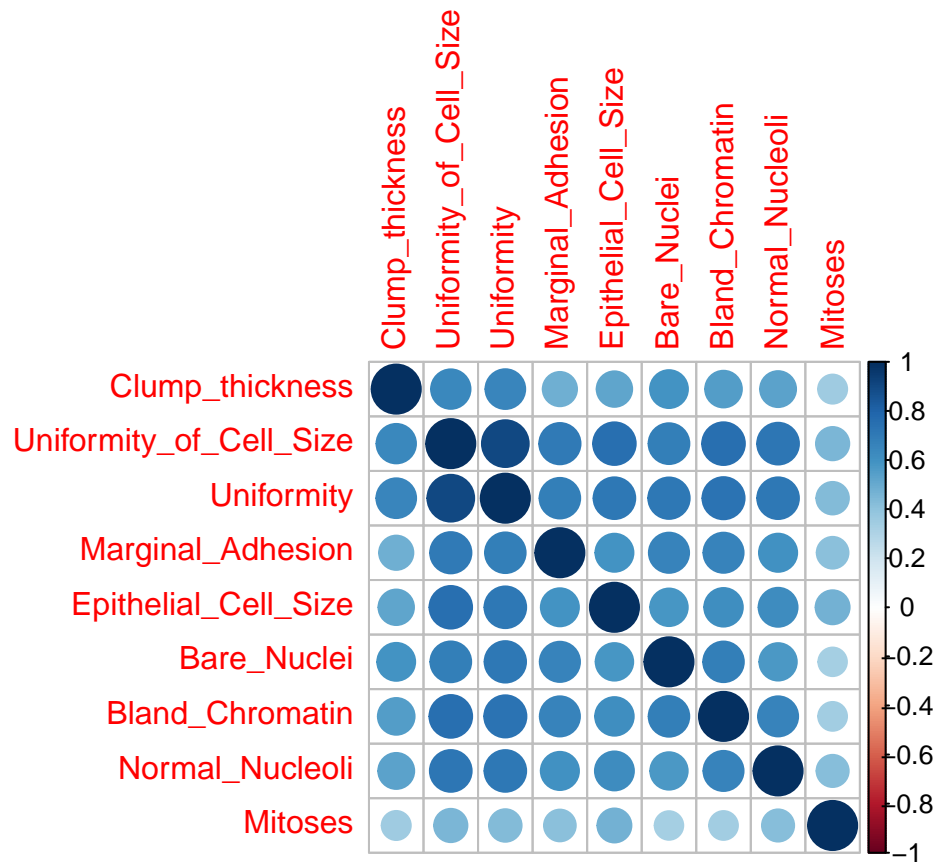
```
Breast_Cancer = Breast_Cancer %>% replace_with_na_all(condition = ~.x == '?')
```

```
Breast_Cancer <- knn.impute(as.matrix(Breast_Cancer), k = 10, cat.var = 2:ncol(Breast_Cancer) - 2,
  to.impute = 1:nrow(Breast_Cancer), using = 1:nrow(Breast_Cancer))
```

```
Breast_Cancer <- data.frame(Breast_Cancer)
```

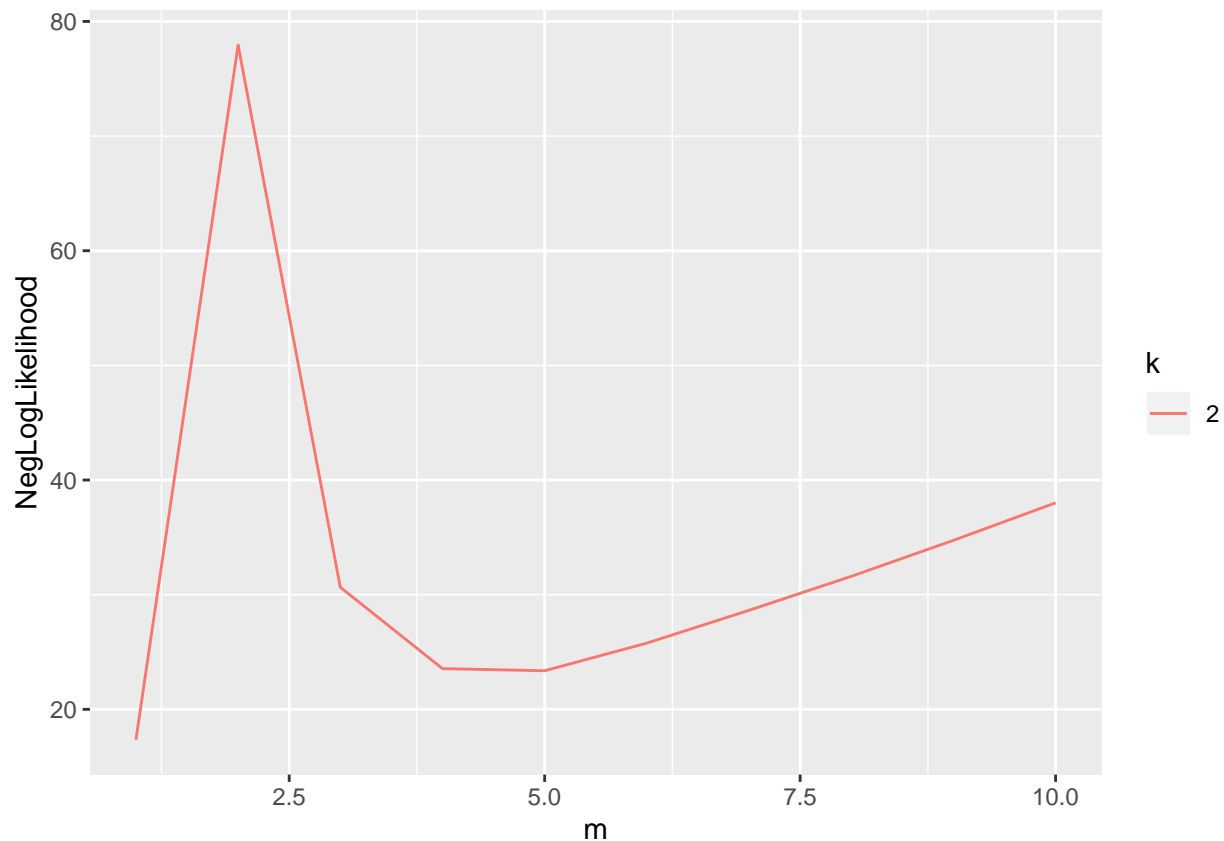
```
Breast_Cancer$Class_cancer = ifelse(Breast_Cancer$Class_cancer == 4, 1,0)
Breast_Cancer = Breast_Cancer[,2:11]
```

```
x = model.matrix(Class_cancer ~ ., Breast_Cancer) [, -1]
corrplot(cor(x))
```



```
lpca = cv.lpca(Breast_Cancer[:,1:9], ks = 2, ms = 1:10)
```

```
plot(lpca)
```



```
cancer.pca <- prcomp(Breast_Cancer[,1:9],center = TRUE , scale. = TRUE)
```

```
summary(cancer.pca)
```

```
## Importance of components:
```

```
##          PC1      PC2      PC3      PC4      PC5      PC6
## Standard deviation  2.4281 0.88207 0.73405 0.68053 0.61598 0.55020
## Proportion of Variance 0.6551 0.08645 0.05987 0.05146 0.04216 0.03364
## Cumulative Proportion 0.6551 0.74152 0.80139 0.85285 0.89501 0.92865
##          PC7      PC8      PC9
## Standard deviation  0.54056 0.51120 0.29776
## Proportion of Variance 0.03247 0.02904 0.00985
## Cumulative Proportion 0.96111 0.99015 1.00000
```

```
cancer.pca[2]
```

```
## $rotation
```

```
##          PC1      PC2      PC3      PC4
## Clump_thickness -0.3028066 -0.14516611 0.862070604 -0.09166947
## Uniformity_of_Cell_Size -0.3809552 -0.04603023 -0.019826189 0.20729219
## Uniformity -0.3774114 -0.08239072 0.033951112 0.17974611
## Marginal_Adhesion -0.3326679 -0.04395983 -0.420781602 -0.47981882
## Epithelial_Cell_Size -0.3360462 0.16741772 -0.110861278 0.39274433
## Bare_Nuclei -0.3346784 -0.25790766 0.005883623 -0.52137706
## Bland_Chromatin -0.3461255 -0.22947889 -0.197768757 -0.02336121
```

## Normal_Nucleoli	-0.3357904	0.02690694	-0.132698092	0.44309157
## Mitoses	-0.2293016	0.90562560	0.095839741	-0.25126451
##	PC5	PC6	PC7	PC8
## Clump_thickness	0.058502664	-0.25726283	-0.05374613	-0.25133759
## Uniformity_of_Cell_Size	-0.136295280	-0.05631810	-0.26026857	0.42365740
## Uniformity	-0.101511676	-0.01019422	-0.17590985	0.57747727
## Marginal_Adhesion	-0.001782729	-0.67379103	-0.06893763	-0.13807397
## Epithelial_Cell_Size	-0.667156855	0.01744502	0.22703887	-0.43703482
## Bare_Nuclei	-0.112981509	0.45755288	0.54641605	0.15104594
## Bland_Chromatin	0.245374357	0.47506251	-0.56001533	-0.42967114
## Normal_Nucleoli	0.658420753	-0.10636780	0.47219743	-0.06657826
## Mitoses	0.126900286	0.17288249	-0.08348781	0.04399019
##	PC9			
## Clump_thickness	-4.607015e-05			
## Uniformity_of_Cell_Size	-7.351311e-01			
## Uniformity	6.651904e-01			
## Marginal_Adhesion	4.797879e-02			
## Epithelial_Cell_Size	6.750739e-02			
## Bare_Nuclei	-7.810562e-02			
## Bland_Chromatin	6.124449e-02			
## Normal_Nucleoli	-1.862371e-02			
## Mitoses	6.823082e-03			