Predicting Breast Cancer

Muiz Rahemtullah

9/11/2019

In this project, we will use machine learning algorithms to predict a binary outcome, in this case whether a woman either has a benign or malignant tumor. We will get a dataset from the UC Irvine Machine Learning database. The rows are different woman and the rows are different features of intrest.

```
uciwd ="https://archive.ics.uci.edu/ml/machine-learning-databases/"
mldata = paste(uciwd,"breast-cancer-wisconsin/breast-cancer-wisconsin.data", sep="")
bcancer = read.csv(mldata)   # Treat the first row as variable names
bcancer
```

```
##      X1000025 X5 X1 X1.1 X1.2 X2 X1.3 X3 X1.4 X1.5 X2.1
## 1     1002945  5  4    4    5  7   10  3    2    1    2
## 2     1015425  3  1    1    1  2    2  3    1    1    2
## 3     1016277  6  8    8    1  3    4  3    7    1    2
## 4     1017023  4  1    1    3  2    1  3    1    1    2
## 5     1017122  8 10   10    8  7   10  9    7    1    4
## 6     1018099  1  1    1    1  2   10  3    1    1    2
## 7     1018561  2  1    2    1  2    1  3    1    1    2
## 8     1033078  2  1    1    1  2    1  1    1    5    2
## 9     1033078  4  2    1    1  2    1  2    1    1    2
## 10    1035283  1  1    1    1  1    1  3    1    1    2
## 11    1036172  2  1    1    1  2    1  2    1    1    2
## 12    1041801  5  3    3    3  2    3  4    4    1    4
## 13    1043999  1  1    1    1  2    3  3    1    1    2
## 14    1044572  8  7    5   10  7    9  5    5    4    4
## 15    1047630  7  4    6    4  6    1  4    3    1    4
## 16    1048672  4  1    1    1  2    1  2    1    1    2
## 17    1049815  4  1    1    1  2    1  3    1    1    2
## 18    1050670 10  7    7    6  4   10  4    1    2    4
## 19    1050718  6  1    1    1  2    1  3    1    1    2
## 20    1054590  7  3    2   10  5   10  5    4    4    4
## 21    1054593 10  5    5    3  6    7  7   10    1    4
## 22    1056784  3  1    1    1  2    1  2    1    1    2
## 23    1057013  8  4    5    1  2    ?  7    3    1    4
## 24    1059552  1  1    1    1  2    1  3    1    1    2
## 25    1065726  5  2    3    4  2    7  3    6    1    4
## 26    1066373  3  2    1    1  1    1  2    1    1    2
## 27    1066979  5  1    1    1  2    1  2    1    1    2
## 28    1067444  2  1    1    1  2    1  2    1    1    2
## 29    1070935  1  1    3    1  2    1  1    1    1    2
## 30    1070935  3  1    1    1  1    1  2    1    1    2
## 31    1071760  2  1    1    1  2    1  3    1    1    2
## 32    1072179 10  7    7    3  8    5  7    4    3    4
## 33    1074610  2  1    1    2  2    1  3    1    1    2
## 34    1075123  3  1    2    1  2    1  2    1    1    2
## 35    1079304  2  1    1    1  2    1  2    1    1    2
## 36    1080185 10 10   10    8  6    1  8    9    1    4
## 37    1081791  6  2    1    1  1    1  7    1    1    2
## 38    1084584  5  4    4    9  2   10  5    6    1    4
```

```
## 39    1091262  2  5    3    3  6    7  7    5    1    4
## 40    1096800  6  6    6    9  6    ?  7    8    1    2
## 41    1099510 10  4    3    1  3    3  6    5    2    4
## 42    1100524  6 10   10    2  8   10  7    3    3    4
## 43    1102573  5  6    5    6 10    1  3    1    1    4
## 44    1103608 10 10   10    4  8    1  8   10    1    4
## 45    1103722  1  1    1    1  2    1  2    1    2    2
## 46    1105257  3  7    7    4  4    9  4    8    1    4
## 47    1105524  1  1    1    1  2    1  2    1    1    2
## 48    1106095  4  1    1    3  2    1  3    1    1    2
## 49    1106829  7  8    7    2  4    8  3    8    2    4
## 50    1108370  9  5    8    1  2    3  2    1    5    4
## 51    1108449  5  3    3    4  2    4  3    4    1    4
## 52    1110102 10  3    6    2  3    5  4   10    2    4
## 53    1110503  5  5    5    8 10    8  7    3    7    4
## 54    1110524 10  5    5    6  8    8  7    1    1    4
## 55    1111249 10  6    6    3  4    5  3    6    1    4
## 56    1112209  8 10   10    1  3    6  3    9    1    4
## 57    1113038  8  2    4    1  5    1  5    4    4    4
## 58    1113483  5  2    3    1  6   10  5    1    1    4
## 59    1113906  9  5    5    2  2    2  5    1    1    4
## 60    1115282  5  3    5    5  3    3  4   10    1    4
## 61    1115293  1  1    1    1  2    2  2    1    1    2
## 62    1116116  9 10   10    1 10    8  3    3    1    4
## 63    1116132  6  3    4    1  5    2  3    9    1    4
## 64    1116192  1  1    1    1  2    1  2    1    1    2
## 65    1116998 10  4    2    1  3    2  4    3   10    4
## 66    1117152  4  1    1    1  2    1  3    1    1    2
## 67    1118039  5  3    4    1  8   10  4    9    1    4
## 68    1120559  8  3    8    3  4    9  8    9    8    4
## 69    1121732  1  1    1    1  2    1  3    2    1    2
## 70    1121919  5  1    3    1  2    1  2    1    1    2
## 71    1123061  6 10    2    8 10    2  7    8   10    4
## 72    1124651  1  3    3    2  2    1  7    2    1    2
## 73    1125035  9  4    5   10  6   10  4    8    1    4
## 74    1126417 10  6    4    1  3    4  3    2    3    4
## 75    1131294  1  1    2    1  2    2  4    2    1    2
## 76    1132347  1  1    4    1  2    1  2    1    1    2
## 77    1133041  5  3    1    2  2    1  2    1    1    2
## 78    1133136  3  1    1    1  2    3  3    1    1    2
## 79    1136142  2  1    1    1  3    1  2    1    1    2
## 80    1137156  2  2    2    1  1    1  7    1    1    2
## 81    1143978  4  1    1    2  2    1  2    1    1    2
## 82    1143978  5  2    1    1  2    1  3    1    1    2
## 83    1147044  3  1    1    1  2    2  7    1    1    2
## 84    1147699  3  5    7    8  8    9  7   10    7    4
## 85    1147748  5 10    6    1 10    4  4   10   10    4
## 86    1148278  3  3    6    4  5    8  4    4    1    4
## 87    1148873  3  6    6    6  5   10  6    8    3    4
## 88    1152331  4  1    1    1  2    1  3    1    1    2
## 89    1155546  2  1    1    2  3    1  2    1    1    2
## 90    1156272  1  1    1    1  2    1  3    1    1    2
## 91    1156948  3  1    1    2  2    1  1    1    1    2
## 92    1157734  4  1    1    1  2    1  3    1    1    2
```

```
## 93    1158247  1  1    1    1  2    1  2    1    1    2
## 94    1160476  2  1    1    1  2    1  3    1    1    2
## 95    1164066  1  1    1    1  2    1  3    1    1    2
## 96    1165297  2  1    1    2  2    1  1    1    1    2
## 97    1165790  5  1    1    1  2    1  3    1    1    2
## 98    1165926  9  6    9    2 10    6  2    9   10    4
## 99    1166630  7  5    6   10  5   10  7    9    4    4
## 100   1166654 10  3    5    1 10    5  3   10    2    4
## 101   1167439  2  3    4    4  2    5  2    5    1    4
## 102   1167471  4  1    2    1  2    1  3    1    1    2
## 103   1168359  8  2    3    1  6    3  7    1    1    4
## 104   1168736 10 10   10   10 10    1  8    8    8    4
## 105   1169049  7  3    4    4  3    3  3    2    7    4
## 106   1170419 10 10   10    8  2   10  4    1    1    4
## 107   1170420  1  6    8   10  8   10  5    7    1    4
## 108   1171710  1  1    1    1  2    1  2    3    1    2
## 109   1171710  6  5    4    4  3    9  7    8    3    4
## 110   1171795  1  3    1    2  2    2  5    3    2    2
## 111   1171845  8  6    4    3  5    9  3    1    1    4
## 112   1172152 10  3    3   10  2   10  7    3    3    4
## 113   1173216 10 10   10    3 10    8  8    1    1    4
## 114   1173235  3  3    2    1  2    3  3    1    1    2
## 115   1173347  1  1    1    1  2    5  1    1    1    2
## 116   1173347  8  3    3    1  2    2  3    2    1    2
## 117   1173509  4  5    5   10  4   10  7    5    8    4
## 118   1173514  1  1    1    1  4    3  1    1    1    2
## 119   1173681  3  2    1    1  2    2  3    1    1    2
## 120   1174057  1  1    2    2  2    1  3    1    1    2
## 121   1174057  4  2    1    1  2    2  3    1    1    2
## 122   1174131 10 10   10    2 10   10  5    3    3    4
## 123   1174428  5  3    5    1  8   10  5    3    1    4
## 124   1175937  5  4    6    7  9    7  8   10    1    4
## 125   1176406  1  1    1    1  2    1  2    1    1    2
## 126   1176881  7  5    3    7  4   10  7    5    5    4
## 127   1177027  3  1    1    1  2    1  3    1    1    2
## 128   1177399  8  3    5    4  5   10  1    6    2    4
## 129   1177512  1  1    1    1 10    1  1    1    1    2
## 130   1178580  5  1    3    1  2    1  2    1    1    2
## 131   1179818  2  1    1    1  2    1  3    1    1    2
## 132   1180194  5 10    8   10  8   10  3    6    3    4
## 133   1180523  3  1    1    1  2    1  2    2    1    2
## 134   1180831  3  1    1    1  3    1  2    1    1    2
## 135   1181356  5  1    1    1  2    2  3    3    1    2
## 136   1182404  4  1    1    1  2    1  2    1    1    2
## 137   1182410  3  1    1    1  2    1  1    1    1    2
## 138   1183240  4  1    2    1  2    1  2    1    1    2
## 139   1183246  1  1    1    1  1    ?  2    1    1    2
## 140   1183516  3  1    1    1  2    1  1    1    1    2
## 141   1183911  2  1    1    1  2    1  1    1    1    2
## 142   1183983  9  5    5    4  4    5  4    3    3    4
## 143   1184184  1  1    1    1  2    5  1    1    1    2
## 144   1184241  2  1    1    1  2    1  2    1    1    2
## 145   1184840  1  1    3    1  2    ?  2    1    1    2
## 146   1185609  3  4    5    2  6    8  4    1    1    4
```

```
## 147   1185610  1  1   1   1  3   2  2   1   1   2
## 148   1187457  3  1   1   3  8   1  5   8   1   2
## 149   1187805  8  8   7   4 10  10  7   8   7   4
## 150   1188472  1  1   1   1  1   1  3   1   1   2
## 151   1189266  7  2   4   1  6  10  5   4   3   4
## 152   1189286 10 10   8   6  4   5  8  10   1   4
## 153   1190394  4  1   1   1  2   3  1   1   1   2
## 154   1190485  1  1   1   1  2   1  1   1   1   2
## 155   1192325  5  5   5   6  3  10  3   1   1   4
## 156   1193091  1  2   2   1  2   1  2   1   1   2
## 157   1193210  2  1   1   1  2   1  3   1   1   2
## 158   1193683  1  1   2   1  3   ?  1   1   1   2
## 159   1196295  9  9  10   3  6  10  7  10   6   4
## 160   1196915 10  7   7   4  5  10  5   7   2   4
## 161   1197080  4  1   1   1  2   1  3   2   1   2
## 162   1197270  3  1   1   1  2   1  3   1   1   2
## 163   1197440  1  1   1   2  1   3  1   1   7   2
## 164   1197510  5  1   1   1  2   ?  3   1   1   2
## 165   1197979  4  1   1   1  2   2  3   2   1   2
## 166   1197993  5  6   7   8  8  10  3  10   3   4
## 167   1198128 10  8  10  10  6   1  3   1  10   4
## 168   1198641  3  1   1   1  2   1  3   1   1   2
## 169   1199219  1  1   1   2  1   1  1   1   1   2
## 170   1199731  3  1   1   1  2   1  1   1   1   2
## 171   1199983  1  1   1   1  2   1  3   1   1   2
## 172   1200772  1  1   1   1  2   1  2   1   1   2
## 173   1200847  6 10  10  10  8  10 10  10   7   4
## 174   1200892  8  6   5   4  3  10  6   1   1   4
## 175   1200952  5  8   7   7 10  10  5   7   1   4
## 176   1201834  2  1   1   1  2   1  3   1   1   2
## 177   1201936  5 10  10   3  8   1  5  10   3   4
## 178   1202125  4  1   1   1  2   1  3   1   1   2
## 179   1202812  5  3   3   3  6  10  3   1   1   4
## 180   1203096  1  1   1   1  1   1  3   1   1   2
## 181   1204242  1  1   1   1  2   1  1   1   1   2
## 182   1204898  6  1   1   1  2   1  3   1   1   2
## 183   1205138  5  8   8   8  5  10  7   8   1   4
## 184   1205579  8  7   6   4  4  10  5   1   1   4
## 185   1206089  2  1   1   1  1   1  3   1   1   2
## 186   1206695  1  5   8   6  5   8  7  10   1   4
## 187   1206841 10  5   6  10  6  10  7   7  10   4
## 188   1207986  5  8   4  10  5   8  9  10   1   4
## 189   1208301  1  2   3   1  2   1  3   1   1   2
## 190   1210963 10 10  10   8  6   8  7  10   1   4
## 191   1211202  7  5  10  10 10  10  4  10   3   4
## 192   1212232  5  1   1   1  2   1  2   1   1   2
## 193   1212251  1  1   1   1  2   1  3   1   1   2
## 194   1212422  3  1   1   1  2   1  3   1   1   2
## 195   1212422  4  1   1   1  2   1  3   1   1   2
## 196   1213375  8  4   4   5  4   7  7   8   2   2
## 197   1213383  5  1   1   4  2   1  3   1   1   2
## 198   1214092  1  1   1   1  2   1  1   1   1   2
## 199   1214556  3  1   1   1  2   1  2   1   1   2
## 200   1214966  9  7   7   5  5  10  7   8   3   4
```

```
## 201  1216694 10  8     8     4 10    10  8     1     1     4
## 202  1216947  1  1     1     1  2     1  3     1     1     2
## 203  1217051  5  1     1     1  2     1  3     1     1     2
## 204  1217264  1  1     1     1  2     1  3     1     1     2
## 205  1218105  5 10    10     9  6    10  7    10     5     4
## 206  1218741 10 10     9     3  7     5  3     5     1     4
## 207  1218860  1  1     1     1  1     1  3     1     1     2
## 208  1218860  1  1     1     1  1     1  3     1     1     2
## 209  1219406  5  1     1     1  1     1  3     1     1     2
## 210  1219525  8 10    10    10  5    10  8    10     6     4
## 211  1219859  8 10     8     8  4     8  7     7     1     4
## 212  1220330  1  1     1     1  2     1  3     1     1     2
## 213  1221863 10 10    10    10  7    10  7    10     4     4
## 214  1222047 10 10    10    10  3    10 10     6     1     4
## 215  1222936  8  7     8     7  5     5  5    10     2     4
## 216  1223282  1  1     1     1  2     1  2     1     1     2
## 217  1223426  1  1     1     1  2     1  3     1     1     2
## 218  1223793  6 10     7     7  6     4  8    10     2     4
## 219  1223967  6  1     3     1  2     1  3     1     1     2
## 220  1224329  1  1     1     2  2     1  3     1     1     2
## 221  1225799 10  6     4     3 10    10  9    10     1     4
## 222  1226012  4  1     1     3  1     5  2     1     1     4
## 223  1226612  7  5     6     3  3     8  7     4     1     4
## 224  1227210 10  5     5     6  3    10  7     9     2     4
## 225  1227244  1  1     1     1  2     1  2     1     1     2
## 226  1227481 10  5     7     4  4    10  8     9     1     4
## 227  1228152  8  9     9     5  3     5  7     7     1     4
## 228  1228311  1  1     1     1  1     1  3     1     1     2
## 229  1230175 10 10    10     3 10    10  9    10     1     4
## 230  1230688  7  4     7     4  3     7  7     6     1     4
## 231  1231387  6  8     7     5  6     8  8     9     2     4
## 232  1231706  8  4     6     3  3     1  4     3     1     2
## 233  1232225 10  4     5     5  5    10  4     1     1     4
## 234  1236043  3  3     2     1  3     1  3     6     1     2
## 235  1241232  3  1     4     1  2     ?  3     1     1     2
## 236  1241559 10  8     8     2  8    10  4     8    10     4
## 237  1241679  9  8     8     5  6     2  4    10     4     4
## 238  1242364  8 10    10     8  6     9  3    10    10     4
## 239  1243256 10  4     3     2  3    10  5     3     2     4
## 240  1270479  5  1     3     3  2     2  2     3     1     2
## 241  1276091  3  1     1     3  1     1  3     1     1     2
## 242  1277018  2  1     1     1  2     1  3     1     1     2
## 243   128059  1  1     1     1  2     5  5     1     1     2
## 244  1285531  1  1     1     1  2     1  3     1     1     2
## 245  1287775  5  1     1     2  2     2  3     1     1     2
## 246   144888  8 10    10     8  5    10  7     8     1     4
## 247   145447  8  4     4     1  2     9  3     3     1     4
## 248   167528  4  1     1     1  2     1  3     6     1     2
## 249   169356  3  1     1     1  2     ?  3     1     1     2
## 250   183913  1  2     2     1  2     1  1     1     1     2
## 251   191250 10  4     4    10  2    10  5     3     3     4
## 252  1017023  6  3     3     5  3    10  3     5     3     2
## 253  1100524  6 10    10     2  8    10  7     3     3     4
## 254  1116116  9 10    10     1 10     8  3     3     1     4
```

```
## 255  1168736  5  6    6    2  4    10  3    6    1    4
## 256  1182404  3  1    1    1  2     1  1    1    1    2
## 257  1182404  3  1    1    1  2     1  2    1    1    2
## 258  1198641  3  1    1    1  2     1  3    1    1    2
## 259   242970  5  7    7    1  5     8  3    4    1    2
## 260   255644 10  5    8   10  3    10  5    1    3    4
## 261   263538  5 10   10    6 10    10 10    6    5    4
## 262   274137  8  8    9    4  5    10  7    8    1    4
## 263   303213 10  4    4   10  6    10  5    5    1    4
## 264   314428  7  9    4   10 10     3  5    3    3    4
## 265  1182404  5  1    4    1  2     1  3    2    1    2
## 266  1198641 10 10    6    3  3    10  4    3    2    4
## 267   320675  3  3    5    2  3    10  7    1    1    4
## 268   324427 10  8    8    2  3     4  8    7    8    4
## 269   385103  1  1    1    1  2     1  3    1    1    2
## 270   390840  8  4    7    1  3    10  3    9    2    4
## 271   411453  5  1    1    1  2     1  3    1    1    2
## 272   320675  3  3    5    2  3    10  7    1    1    4
## 273   428903  7  2    4    1  3     4  3    3    1    4
## 274   431495  3  1    1    1  2     1  3    2    1    2
## 275   432809  3  1    3    1  2     ?  2    1    1    2
## 276   434518  3  1    1    1  2     1  2    1    1    2
## 277   452264  1  1    1    1  2     1  2    1    1    2
## 278   456282  1  1    1    1  2     1  3    1    1    2
## 279   476903 10  5    7    3  3     7  3    3    8    4
## 280   486283  3  1    1    1  2     1  3    1    1    2
## 281   486662  2  1    1    2  2     1  3    1    1    2
## 282   488173  1  4    3   10  4    10  5    6    1    4
## 283   492268 10  4    6    1  2    10  5    3    1    4
## 284   508234  7  4    5   10  2    10  3    8    2    4
## 285   527363  8 10   10   10  8    10 10    7    3    4
## 286   529329 10 10   10   10 10    10  4   10   10    4
## 287   535331  3  1    1    1  3     1  2    1    1    2
## 288   543558  6  1    3    1  4     5  5   10    1    4
## 289   555977  5  6    6    8  6    10  4   10    4    4
## 290   560680  1  1    1    1  2     1  1    1    1    2
## 291   561477  1  1    1    1  2     1  3    1    1    2
## 292   563649  8  8    8    1  2     ?  6   10    1    4
## 293   601265 10  4    4    6  2    10  2    3    1    4
## 294   606140  1  1    1    1  2     ?  2    1    1    2
## 295   606722  5  5    7    8  6    10  7    4    1    4
## 296   616240  5  3    4    3  4     5  4    7    1    2
## 297    61634  5  4    3    1  2     ?  2    3    1    2
## 298   625201  8  2    1    1  5     1  1    1    1    2
## 299    63375  9  1    2    6  4    10  7    7    2    4
## 300   635844  8  4   10    5  4     4  7   10    1    4
## 301   636130  1  1    1    1  2     1  3    1    1    2
## 302   640744 10 10   10    7  9    10  7   10   10    4
## 303   646904  1  1    1    1  2     1  3    1    1    2
## 304   653777  8  3    4    9  3    10  3    3    1    4
## 305   659642 10  8    4    4  4    10  3   10    4    4
## 306   666090  1  1    1    1  2     1  3    1    1    2
## 307   666942  1  1    1    1  2     1  3    1    1    2
## 308   667204  7  8    7    6  4     3  8    8    4    4
```

```
## 309   673637  3  1   1   1  2   5  5   1   1   2
## 310   684955  2  1   1   1  3   1  2   1   1   2
## 311   688033  1  1   1   1  2   1  1   1   1   2
## 312   691628  8  6   4  10 10   1  3   5   1   4
## 313   693702  1  1   1   1  2   1  1   1   1   2
## 314   704097  1  1   1   1  1   1  2   1   1   2
## 315   704168  4  6   5   6  7   ?  4   9   1   2
## 316   706426  5  5   5   2  5  10  4   3   1   4
## 317   709287  6  8   7   8  6   8  8   9   1   4
## 318   718641  1  1   1   1  5   1  3   1   1   2
## 319   721482  4  4   4   4  6   5  7   3   1   2
## 320   730881  7  6   3   2  5  10  7   4   6   4
## 321   733639  3  1   1   1  2   ?  3   1   1   2
## 322   733639  3  1   1   1  2   1  3   1   1   2
## 323   733823  5  4   6  10  2  10  4   1   1   4
## 324   740492  1  1   1   1  2   1  3   1   1   2
## 325   743348  3  2   2   1  2   1  2   3   1   2
## 326   752904 10  1   1   1  2  10  5   4   1   4
## 327   756136  1  1   1   1  2   1  2   1   1   2
## 328   760001  8 10   3   2  6   4  3  10   1   4
## 329   760239 10  4   6   4  5  10  7   1   1   4
## 330    76389 10  4   7   2  2   8  6   1   1   4
## 331   764974  5  1   1   1  2   1  3   1   2   2
## 332   770066  5  2   2   2  2   1  2   2   1   2
## 333   785208  5  4   6   6  4  10  4   3   1   4
## 334   785615  8  6   7   3  3  10  3   4   2   4
## 335   792744  1  1   1   1  2   1  1   1   1   2
## 336   797327  6  5   5   8  4  10  3   4   1   4
## 337   798429  1  1   1   1  2   1  3   1   1   2
## 338   704097  1  1   1   1  1   1  2   1   1   2
## 339   806423  8  5   5   5  2  10  4   3   1   4
## 340   809912 10  3   3   1  2  10  7   6   1   4
## 341   810104  1  1   1   1  2   1  3   1   1   2
## 342   814265  2  1   1   1  2   1  1   1   1   2
## 343   814911  1  1   1   1  2   1  1   1   1   2
## 344   822829  7  6   4   8 10  10  9   5   3   4
## 345   826923  1  1   1   1  2   1  1   1   1   2
## 346   830690  5  2   2   2  3   1  1   3   1   2
## 347   831268  1  1   1   1  1   1  1   3   1   2
## 348   832226  3  4   4  10  5   1  3   3   1   4
## 349   832567  4  2   3   5  3   8  7   6   1   4
## 350   836433  5  1   1   3  2   1  1   1   1   2
## 351   837082  2  1   1   1  2   1  3   1   1   2
## 352   846832  3  4   5   3  7   3  4   6   1   2
## 353   850831  2  7  10  10  7  10  4   9   4   4
## 354   855524  1  1   1   1  2   1  2   1   1   2
## 355   857774  4  1   1   1  3   1  2   2   1   2
## 356   859164  5  3   3   1  3   3  3   3   3   4
## 357   859350  8 10  10   7 10  10  7   3   8   4
## 358   866325  8 10   5   3  8   4  4  10   3   4
## 359   873549 10  3   5   4  3   7  3   5   3   4
## 360   877291  6 10  10  10 10  10  8  10  10   4
## 361   877943  3 10   3  10  6  10  5   1   4   4
## 362   888169  3  2   2   1  4   3  2   1   1   2
```

```
## 363   888523  4  4    4    2  2    3  2    1    1    2
## 364   896404  2  1    1    1  2    1  3    1    1    2
## 365   897172  2  1    1    1  2    1  2    1    1    2
## 366    95719  6 10   10   10  8   10  7   10    7    4
## 367   160296  5  8    8   10  5   10  8   10    3    4
## 368   342245  1  1    3    1  2    1  1    1    1    2
## 369   428598  1  1    3    1  1    1  2    1    1    2
## 370   492561  4  3    2    1  3    1  2    1    1    2
## 371   493452  1  1    3    1  2    1  1    1    1    2
## 372   493452  4  1    2    1  2    1  2    1    1    2
## 373   521441  5  1    1    2  2    1  2    1    1    2
## 374   560680  3  1    2    1  2    1  2    1    1    2
## 375   636437  1  1    1    1  2    1  1    1    1    2
## 376   640712  1  1    1    1  2    1  2    1    1    2
## 377   654244  1  1    1    1  1    1  2    1    1    2
## 378   657753  3  1    1    4  3    1  2    2    1    2
## 379   685977  5  3    4    1  4    1  3    1    1    2
## 380   805448  1  1    1    1  2    1  1    1    1    2
## 381   846423 10  6    3    6  4   10  7    8    4    4
## 382  1002504  3  2    2    2  2    1  3    2    1    2
## 383  1022257  2  1    1    1  2    1  1    1    1    2
## 384  1026122  2  1    1    1  2    1  1    1    1    2
## 385  1071084  3  3    2    2  3    1  1    2    3    2
## 386  1080233  7  6    6    3  2   10  7    1    1    4
## 387  1114570  5  3    3    2  3    1  3    1    1    2
## 388  1114570  2  1    1    1  2    1  2    2    1    2
## 389  1116715  5  1    1    1  3    2  2    2    1    2
## 390  1131411  1  1    1    2  2    1  2    1    1    2
## 391  1151734 10  8    7    4  3   10  7    9    1    4
## 392  1156017  3  1    1    1  2    1  2    1    1    2
## 393  1158247  1  1    1    1  1    1  1    1    1    2
## 394  1158405  1  2    3    1  2    1  2    1    1    2
## 395  1168278  3  1    1    1  2    1  2    1    1    2
## 396  1176187  3  1    1    1  2    1  3    1    1    2
## 397  1196263  4  1    1    1  2    1  1    1    1    2
## 398  1196475  3  2    1    1  2    1  2    2    1    2
## 399  1206314  1  2    3    1  2    1  1    1    1    2
## 400  1211265  3 10    8    7  6    9  9    3    8    4
## 401  1213784  3  1    1    1  2    1  1    1    1    2
## 402  1223003  5  3    3    1  2    1  2    1    1    2
## 403  1223306  3  1    1    1  2    4  1    1    1    2
## 404  1223543  1  2    1    3  2    1  1    2    1    2
## 405  1229929  1  1    1    1  2    1  2    1    1    2
## 406  1231853  4  2    2    1  2    1  2    1    1    2
## 407  1234554  1  1    1    1  2    1  2    1    1    2
## 408  1236837  2  3    2    2  2    2  3    1    1    2
## 409  1237674  3  1    2    1  2    1  2    1    1    2
## 410  1238021  1  1    1    1  2    1  2    1    1    2
## 411  1238464  1  1    1    1  1    ?  2    1    1    2
## 412  1238633 10 10   10    6  8    4  8    5    1    4
## 413  1238915  5  1    2    1  2    1  3    1    1    2
## 414  1238948  8  5    6    2  3   10  6    6    1    4
## 415  1239232  3  3    2    6  3    3  3    5    1    2
## 416  1239347  8  7    8    5 10   10  7    2    1    4
```

```
## 417   1239967  1  1   1   1  2   1  2   1   1   2
## 418   1240337  5  2   2   2  2   2  3   2   2   2
## 419   1253505  2  3   1   1  5   1  1   1   1   2
## 420   1255384  3  2   2   3  2   3  3   1   1   2
## 421   1257200 10 10  10   7 10  10  8   2   1   4
## 422   1257648  4  3   3   1  2   1  3   3   1   2
## 423   1257815  5  1   3   1  2   1  2   1   1   2
## 424   1257938  3  1   1   1  2   1  1   1   1   2
## 425   1258549  9 10  10  10 10  10 10  10   1   4
## 426   1258556  5  3   6   1  2   1  1   1   1   2
## 427   1266154  8  7   8   2  4   2  5  10   1   4
## 428   1272039  1  1   1   1  2   1  2   1   1   2
## 429   1276091  2  1   1   1  2   1  2   1   1   2
## 430   1276091  1  3   1   1  2   1  2   2   1   2
## 431   1276091  5  1   1   3  4   1  3   2   1   2
## 432   1277629  5  1   1   1  2   1  2   2   1   2
## 433   1293439  3  2   2   3  2   1  1   1   1   2
## 434   1293439  6  9   7   5  5   8  4   2   1   2
## 435   1294562 10  8  10   1  3  10  5   1   1   4
## 436   1295186 10 10  10   1  6   1  2   8   1   4
## 437    527337  4  1   1   1  2   1  1   1   1   2
## 438    558538  4  1   3   3  2   1  1   1   1   2
## 439    566509  5  1   1   1  2   1  1   1   1   2
## 440    608157 10  4   3  10  4  10 10   1   1   4
## 441    677910  5  2   2   4  2   4  1   1   1   2
## 442    734111  1  1   1   3  2   3  1   1   1   2
## 443    734111  1  1   1   1  2   2  1   1   1   2
## 444    780555  5  1   1   6  3   1  2   1   1   2
## 445    827627  2  1   1   1  2   1  1   1   1   2
## 446   1049837  1  1   1   1  2   1  1   1   1   2
## 447   1058849  5  1   1   1  2   1  1   1   1   2
## 448   1182404  1  1   1   1  1   1  1   1   1   2
## 449   1193544  5  7   9   8  6  10  8  10   1   4
## 450   1201870  4  1   1   3  1   1  2   1   1   2
## 451   1202253  5  1   1   1  2   1  1   1   1   2
## 452   1227081  3  1   1   3  2   1  1   1   1   2
## 453   1230994  4  5   5   8  6  10 10   7   1   4
## 454   1238410  2  3   1   1  3   1  1   1   1   2
## 455   1246562 10  2   2   1  2   6  1   1   2   4
## 456   1257470 10  6   5   8  5  10  8   6   1   4
## 457   1259008  8  8   9   6  6   3 10  10   1   4
## 458   1266124  5  1   2   1  2   1  1   1   1   2
## 459   1267898  5  1   3   1  2   1  1   1   1   2
## 460   1268313  5  1   1   3  2   1  1   1   1   2
## 461   1268804  3  1   1   1  2   5  1   1   1   2
## 462   1276091  6  1   1   3  2   1  1   1   1   2
## 463   1280258  4  1   1   1  2   1  1   2   1   2
## 464   1293966  4  1   1   1  2   1  1   1   1   2
## 465   1296572 10  9   8   7  6   4  7  10   3   4
## 466   1298416 10  6   6   2  4  10  9   7   1   4
## 467   1299596  6  6   6   5  4  10  7   6   2   4
## 468   1105524  4  1   1   1  2   1  1   1   1   2
## 469   1181685  1  1   2   1  2   1  2   1   1   2
## 470   1211594  3  1   1   1  1   1  2   1   1   2
```

```
## 471  1238777  6  1     1    3  2    1  1    1    1    2
## 472  1257608  6  1     1    1  1    1  1    1    1    2
## 473  1269574  4  1     1    1  2    1  1    1    1    2
## 474  1277145  5  1     1    1  2    1  1    1    1    2
## 475  1287282  3  1     1    1  2    1  1    1    1    2
## 476  1296025  4  1     2    1  2    1  1    1    1    2
## 477  1296263  4  1     1    1  2    1  1    1    1    2
## 478  1296593  5  2     1    1  2    1  1    1    1    2
## 479  1299161  4  8     7   10  4   10  7    5    1    4
## 480  1301945  5  1     1    1  1    1  1    1    1    2
## 481  1302428  5  3     2    4  2    1  1    1    1    2
## 482  1318169  9 10    10   10 10    5 10   10   10    4
## 483   474162  8  7     8    5  5   10  9   10    1    4
## 484   787451  5  1     2    1  2    1  1    1    1    2
## 485  1002025  1  1     1    3  1    3  1    1    1    2
## 486  1070522  3  1     1    1  1    1  2    1    1    2
## 487  1073960 10 10    10   10  6   10  8    1    5    4
## 488  1076352  3  6     4   10  3    3  3    4    1    4
## 489  1084139  6  3     2    1  3    4  4    1    1    4
## 490  1115293  1  1     1    1  2    1  1    1    1    2
## 491  1119189  5  8     9    4  3   10  7    1    1    4
## 492  1133991  4  1     1    1  1    1  2    1    1    2
## 493  1142706  5 10    10   10  6   10  6    5    2    4
## 494  1155967  5  1     2   10  4    5  2    1    1    2
## 495  1170945  3  1     1    1  1    1  2    1    1    2
## 496  1181567  1  1     1    1  1    1  1    1    1    2
## 497  1182404  4  2     1    1  2    1  1    1    1    2
## 498  1204558  4  1     1    1  2    1  2    1    1    2
## 499  1217952  4  1     1    1  2    1  2    1    1    2
## 500  1224565  6  1     1    1  2    1  3    1    1    2
## 501  1238186  4  1     1    1  2    1  2    1    1    2
## 502  1253917  4  1     1    2  2    1  2    1    1    2
## 503  1265899  4  1     1    1  2    1  3    1    1    2
## 504  1268766  1  1     1    1  2    1  1    1    1    2
## 505  1277268  3  3     1    1  2    1  1    1    1    2
## 506  1286943  8 10    10   10  7    5  4    8    7    4
## 507  1295508  1  1     1    1  2    4  1    1    1    2
## 508  1297327  5  1     1    1  2    1  1    1    1    2
## 509  1297522  2  1     1    1  2    1  1    1    1    2
## 510  1298360  1  1     1    1  2    1  1    1    1    2
## 511  1299924  5  1     1    1  2    1  2    1    1    2
## 512  1299994  5  1     1    1  2    1  1    1    1    2
## 513  1304595  3  1     1    1  1    1  2    1    1    2
## 514  1306282  6  6     7   10  3   10  8   10    2    4
## 515  1313325  4 10     4    7  3   10  9   10    1    4
## 516  1320077  1  1     1    1  1    1  1    1    1    2
## 517  1320077  1  1     1    1  1    1  2    1    1    2
## 518  1320304  3  1     2    2  2    1  1    1    1    2
## 519  1330439  4  7     8    3  4   10  9    1    1    4
## 520   333093  1  1     1    1  3    1  1    1    1    2
## 521   369565  4  1     1    1  3    1  1    1    1    2
## 522   412300 10  4     5    4  3    5  7    3    1    4
## 523   672113  7  5     6   10  4   10  5    3    1    4
## 524   749653  3  1     1    1  2    1  2    1    1    2
```

```
## 525   769612  3  1   1   2  2   1  1   1   1   2
## 526   769612  4  1   1   1  2   1  1   1   1   2
## 527   798429  4  1   1   1  2   1  3   1   1   2
## 528   807657  6  1   3   2  2   1  1   1   1   2
## 529  8233704  4  1   1   1  1   1  2   1   1   2
## 530   837480  7  4   4   3  4  10  6   9   1   4
## 531   867392  4  2   2   1  2   1  2   1   1   2
## 532   869828  1  1   1   1  1   1  3   1   1   2
## 533  1043068  3  1   1   1  2   1  2   1   1   2
## 534  1056171  2  1   1   1  2   1  2   1   1   2
## 535  1061990  1  1   3   2  2   1  3   1   1   2
## 536  1113061  5  1   1   1  2   1  3   1   1   2
## 537  1116192  5  1   2   1  2   1  3   1   1   2
## 538  1135090  4  1   1   1  2   1  2   1   1   2
## 539  1145420  6  1   1   1  2   1  2   1   1   2
## 540  1158157  5  1   1   1  2   2  2   1   1   2
## 541  1171578  3  1   1   1  2   1  1   1   1   2
## 542  1174841  5  3   1   1  2   1  1   1   1   2
## 543  1184586  4  1   1   1  2   1  2   1   1   2
## 544  1186936  2  1   3   2  2   1  2   1   1   2
## 545  1197527  5  1   1   1  2   1  2   1   1   2
## 546  1222464  6 10  10  10  4  10  7  10   1   4
## 547  1240603  2  1   1   1  1   1  1   1   1   2
## 548  1240603  3  1   1   1  1   1  1   1   1   2
## 549  1241035  7  8   3   7  4   5  7   8   2   4
## 550  1287971  3  1   1   1  2   1  2   1   1   2
## 551  1289391  1  1   1   1  2   1  3   1   1   2
## 552  1299924  3  2   2   2  2   1  4   2   1   2
## 553  1306339  4  4   2   1  2   5  2   1   2   2
## 554  1313658  3  1   1   1  2   1  1   1   1   2
## 555  1313982  4  3   1   1  2   1  4   8   1   2
## 556  1321264  5  2   2   2  1   1  2   1   1   2
## 557  1321321  5  1   1   3  2   1  1   1   1   2
## 558  1321348  2  1   1   1  2   1  2   1   1   2
## 559  1321931  5  1   1   1  2   1  2   1   1   2
## 560  1321942  5  1   1   1  2   1  3   1   1   2
## 561  1321942  5  1   1   1  2   1  3   1   1   2
## 562  1328331  1  1   1   1  2   1  3   1   1   2
## 563  1328755  3  1   1   1  2   1  2   1   1   2
## 564  1331405  4  1   1   1  2   1  3   2   1   2
## 565  1331412  5  7  10  10  5  10 10  10   1   4
## 566  1333104  3  1   2   1  2   1  3   1   1   2
## 567  1334071  4  1   1   1  2   3  2   1   1   2
## 568  1343068  8  4   4   1  6  10  2   5   2   4
## 569  1343374 10 10   8  10  6   5 10   3   1   4
## 570  1344121  8 10   4   4  8  10  8   2   1   4
## 571   142932  7  6  10   5  3  10  9  10   2   4
## 572   183936  3  1   1   1  2   1  2   1   1   2
## 573   324382  1  1   1   1  2   1  2   1   1   2
## 574   378275 10  9   7   3  4   2  7   7   1   4
## 575   385103  5  1   2   1  2   1  3   1   1   2
## 576   690557  5  1   1   1  2   1  2   1   1   2
## 577   695091  1  1   1   1  2   1  2   1   1   2
## 578   695219  1  1   1   1  2   1  2   1   1   2
```

```
## 579   824249  1  1   1   1  2   1  3   1   1   2
## 580   871549  5  1   2   1  2   1  2   1   1   2
## 581   878358  5  7  10   6  5  10  7   5   1   4
## 582  1107684  6 10   5   5  4  10  6  10   1   4
## 583  1115762  3  1   1   1  2   1  1   1   1   2
## 584  1217717  5  1   1   6  3   1  1   1   1   2
## 585  1239420  1  1   1   1  2   1  1   1   1   2
## 586  1254538  8 10  10  10  6  10 10  10   1   4
## 587  1261751  5  1   1   1  2   1  2   2   1   2
## 588  1268275  9  8   8   9  6   3  4   1   1   4
## 589  1272166  5  1   1   1  2   1  1   1   1   2
## 590  1294261  4 10   8   5  4   1 10   1   1   4
## 591  1295529  2  5   7   6  4  10  7   6   1   4
## 592  1298484 10  3   4   5  3  10  4   1   1   4
## 593  1311875  5  1   2   1  2   1  1   1   1   2
## 594  1315506  4  8   6   3  4  10  7   1   1   4
## 595  1320141  5  1   1   1  2   1  2   1   1   2
## 596  1325309  4  1   2   1  2   1  2   1   1   2
## 597  1333063  5  1   3   1  2   1  3   1   1   2
## 598  1333495  3  1   1   1  2   1  2   1   1   2
## 599  1334659  5  2   4   1  1   1  1   1   1   2
## 600  1336798  3  1   1   1  2   1  2   1   1   2
## 601  1344449  1  1   1   1  1   1  2   1   1   2
## 602  1350568  4  1   1   1  2   1  2   1   1   2
## 603  1352663  5  4   6   8  4   1  8  10   1   4
## 604   188336  5  3   2   8  5  10  8   1   2   4
## 605   352431 10  5  10   3  5   8  7   8   3   4
## 606   353098  4  1   1   2  2   1  1   1   1   2
## 607   411453  1  1   1   1  2   1  1   1   1   2
## 608   557583  5 10  10  10 10  10 10   1   1   4
## 609   636375  5  1   1   1  2   1  1   1   1   2
## 610   736150 10  4   3  10  3  10  7   1   2   4
## 611   803531  5 10  10  10  5   2  8   5   1   4
## 612   822829  8 10  10  10  6  10 10  10  10   4
## 613  1016634  2  3   1   1  2   1  2   1   1   2
## 614  1031608  2  1   1   1  1   1  2   1   1   2
## 615  1041043  4  1   3   1  2   1  2   1   1   2
## 616  1042252  3  1   1   1  2   1  2   1   1   2
## 617  1057067  1  1   1   1  1   ?  1   1   1   2
## 618  1061990  4  1   1   1  2   1  2   1   1   2
## 619  1073836  5  1   1   1  2   1  2   1   1   2
## 620  1083817  3  1   1   1  2   1  2   1   1   2
## 621  1096352  6  3   3   3  3   2  6   1   1   2
## 622  1140597  7  1   2   3  2   1  2   1   1   2
## 623  1149548  1  1   1   1  2   1  1   1   1   2
## 624  1174009  5  1   1   2  1   1  2   1   1   2
## 625  1183596  3  1   3   1  3   4  1   1   1   2
## 626  1190386  4  6   6   5  7   6  7   7   3   4
## 627  1190546  2  1   1   1  2   5  1   1   1   2
## 628  1213273  2  1   1   1  2   1  1   1   1   2
## 629  1218982  4  1   1   1  2   1  1   1   1   2
## 630  1225382  6  2   3   1  2   1  1   1   1   2
## 631  1235807  5  1   1   1  2   1  2   1   1   2
## 632  1238777  1  1   1   1  2   1  1   1   1   2
```

```
## 633  1253955  8  7   4   4  5   3  5  10   1   4
## 634  1257366  3  1   1   1  2   1  1   1   1   2
## 635  1260659  3  1   4   1  2   1  1   1   1   2
## 636  1268952 10 10   7   8  7   1 10  10   3   4
## 637  1275807  4  2   4   3  2   2  2   1   1   2
## 638  1277792  4  1   1   1  2   1  1   1   1   2
## 639  1277792  5  1   1   3  2   1  1   1   1   2
## 640  1285722  4  1   1   3  2   1  1   1   1   2
## 641  1288608  3  1   1   1  2   1  2   1   1   2
## 642  1290203  3  1   1   1  2   1  2   1   1   2
## 643  1294413  1  1   1   1  2   1  1   1   1   2
## 644  1299596  2  1   1   1  2   1  1   1   1   2
## 645  1303489  3  1   1   1  2   1  2   1   1   2
## 646  1311033  1  2   2   1  2   1  1   1   1   2
## 647  1311108  1  1   1   3  2   1  1   1   1   2
## 648  1315807  5 10  10  10 10   2 10  10  10   4
## 649  1318671  3  1   1   1  2   1  2   1   1   2
## 650  1319609  3  1   1   2  3   4  1   1   1   2
## 651  1323477  1  2   1   3  2   1  2   1   1   2
## 652  1324572  5  1   1   1  2   1  2   2   1   2
## 653  1324681  4  1   1   1  2   1  2   1   1   2
## 654  1325159  3  1   1   1  2   1  3   1   1   2
## 655  1326892  3  1   1   1  2   1  2   1   1   2
## 656  1330361  5  1   1   1  2   1  2   1   1   2
## 657  1333877  5  4   5   1  8   1  3   6   1   2
## 658  1334015  7  8   8   7  3  10  7   2   3   4
## 659  1334667  1  1   1   1  2   1  1   1   1   2
## 660  1339781  1  1   1   1  2   1  2   1   1   2
## 661  1339781  4  1   1   1  2   1  3   1   1   2
## 662 13454352  1  1   3   1  2   1  2   1   1   2
## 663  1345452  1  1   3   1  2   1  2   1   1   2
## 664  1345593  3  1   1   3  2   1  2   1   1   2
## 665  1347749  1  1   1   1  2   1  1   1   1   2
## 666  1347943  5  2   2   2  2   1  1   1   2   2
## 667  1348851  3  1   1   1  2   1  3   1   1   2
## 668  1350319  5  7   4   1  6   1  7  10   3   4
## 669  1350423  5 10  10   8  5   5  7  10   1   4
## 670  1352848  3 10   7   8  5   8  7   4   1   4
## 671  1353092  3  2   1   2  2   1  3   1   1   2
## 672  1354840  2  1   1   1  2   1  3   1   1   2
## 673  1354840  5  3   2   1  3   1  1   1   1   2
## 674  1355260  1  1   1   1  2   1  2   1   1   2
## 675  1365075  4  1   4   1  2   1  1   1   1   2
## 676  1365328  1  1   2   1  2   1  2   1   1   2
## 677  1368267  5  1   1   1  2   1  1   1   1   2
## 678  1368273  1  1   1   1  2   1  1   1   1   2
## 679  1368882  2  1   1   1  2   1  1   1   1   2
## 680  1369821 10 10  10  10  5  10 10  10   7   4
## 681  1371026  5 10  10  10  4  10  5   6   3   4
## 682  1371920  5  1   1   1  2   1  3   2   1   2
## 683   466906  1  1   1   1  2   1  1   1   1   2
## 684   466906  1  1   1   1  2   1  1   1   1   2
## 685   534555  1  1   1   1  2   1  1   1   1   2
## 686   536708  1  1   1   1  2   1  1   1   1   2
```

```
## 687   566346  3  1    1    1  2    1  2    3    1    2
## 688   603148  4  1    1    1  2    1  1    1    1    2
## 689   654546  1  1    1    1  2    1  1    1    8    2
## 690   654546  1  1    1    3  2    1  1    1    1    2
## 691   695091  5 10   10    5  4    5  4    4    1    4
## 692   714039  3  1    1    1  2    1  1    1    1    2
## 693   763235  3  1    1    1  2    1  2    1    2    2
## 694   776715  3  1    1    1  3    2  1    1    1    2
## 695   841769  2  1    1    1  2    1  1    1    1    2
## 696   888820  5 10   10    3  7    3  8   10    2    4
## 697   897471  4  8    6    4  3    4 10    6    1    4
## 698   897471  4  8    8    5  4    5 10    4    1    4
```

```
bcancer = read.csv(mldata, header=F) # Treat the data begins from the first row

colnames(bcancer)=c("ID","clump_thick","cell_size","cell_shape", "marginal","epithelial","nuclei",
                    "chromatin","nucleoli","mitoses","class")
str(bcancer)
```

```
## 'data.frame':    699 obs. of  11 variables:
##  $ ID         : int  1000025 1002945 1015425 1016277 1017023 1017122 1018099 1018561 1033078 1033078
##  $ clump_thick: int  5 5 3 6 4 8 1 2 2 4 ...
##  $ cell_size  : int  1 4 1 8 1 10 1 1 1 2 ...
##  $ cell_shape : int  1 4 1 8 1 10 1 2 1 1 ...
##  $ marginal   : int  1 5 1 1 3 8 1 1 1 1 ...
##  $ epithelial : int  2 7 2 3 2 7 2 2 2 2 ...
##  $ nuclei     : Factor w/ 11 levels "?","1","10","2",..: 2 3 4 6 2 3 3 2 2 2 ...
##  $ chromatin  : int  3 3 3 3 3 9 3 3 1 2 ...
##  $ nucleoli   : int  1 2 1 7 1 7 1 1 1 1 ...
##  $ mitoses    : int  1 1 1 1 1 1 1 1 5 1 ...
##  $ class      : int  2 2 2 2 2 4 2 2 2 2 ...
```

```
summary(bcancer)
```

```
##        ID             clump_thick       cell_size        cell_shape
##  Min.   :   61634   Min.   : 1.000   Min.   : 1.000   Min.   : 1.000
##  1st Qu.:  870688   1st Qu.: 2.000   1st Qu.: 1.000   1st Qu.: 1.000
##  Median : 1171710   Median : 4.000   Median : 1.000   Median : 1.000
##  Mean   : 1071704   Mean   : 4.418   Mean   : 3.134   Mean   : 3.207
##  3rd Qu.: 1238298   3rd Qu.: 6.000   3rd Qu.: 5.000   3rd Qu.: 5.000
##  Max.   :13454352   Max.   :10.000   Max.   :10.000   Max.   :10.000
##
##     marginal        epithelial         nuclei      chromatin
##  Min.   : 1.000   Min.   : 1.000   1      :402   Min.   : 1.000
##  1st Qu.: 1.000   1st Qu.: 2.000   10     :132   1st Qu.: 2.000
##  Median : 1.000   Median : 2.000   2      : 30   Median : 3.000
##  Mean   : 2.807   Mean   : 3.216   5      : 30   Mean   : 3.438
##  3rd Qu.: 4.000   3rd Qu.: 4.000   3      : 28   3rd Qu.: 5.000
##  Max.   :10.000   Max.   :10.000   8      : 21   Max.   :10.000
##                                    (Other): 56
##     nucleoli         mitoses           class
##  Min.   : 1.000   Min.   : 1.000   Min.   :2.00
##  1st Qu.: 1.000   1st Qu.: 1.000   1st Qu.:2.00
##  Median : 1.000   Median : 1.000   Median :2.00
##  Mean   : 2.867   Mean   : 1.589   Mean   :2.69
```

```
##  3rd Qu.: 4.000   3rd Qu.: 1.000   3rd Qu.:4.00
##  Max.   :10.000   Max.   :10.000   Max.   :4.00
##
```

```
table(bcancer$nuclei)
```

```
##
##    ?   1  10   2   3   4   5   6   7   8   9
##   16 402 132  30  28  19  30   4   8  21   9
```

```
bcancer$nuclei=as.numeric(gsub("\\?","NA",bcancer$nuclei))
```

```
## Warning: NAs introduced by coercion
```

```
library(Hmisc)
```

```
## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

## Loading required package: ggplot2

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```
bcancer$nuclei <- impute(bcancer$nuclei, mean)
anyNA(bcancer$nuclei)
```

```
## [1] FALSE
```

```
attach(bcancer)
bcancer$class <- as.numeric(gsub(2, 0, bcancer$class))
bcancer$class <- as.numeric(gsub(4, 1, bcancer$class))
bcancer$nuclei <- as.integer(bcancer$nuclei)
str(bcancer)
```

```
## 'data.frame':    699 obs. of  11 variables:
##  $ ID         : int  1000025 1002945 1015425 1016277 1017023 1017122 1018099 1018561 1033078 1033078
##  $ clump_thick: int  5 5 3 6 4 8 1 2 2 4 ...
##  $ cell_size  : int  1 4 1 8 1 10 1 1 1 2 ...
##  $ cell_shape : int  1 4 1 8 1 10 1 2 1 1 ...
##  $ marginal   : int  1 5 1 1 3 8 1 1 1 1 ...
##  $ epithelial : int  2 7 2 3 2 7 2 2 2 2 ...
##  $ nuclei     : int  1 10 2 4 1 10 10 1 1 1 ...
##  $ chromatin  : int  3 3 3 3 3 9 3 3 1 2 ...
##  $ nucleoli   : int  1 2 1 7 1 7 1 1 1 1 ...
##  $ mitoses    : int  1 1 1 1 1 1 1 1 5 1 ...
##  $ class      : num  0 0 0 0 0 1 0 0 0 0 ...
```

Now we will divide the data into the training and testing set.

```
library(caret)
```

```
##
## Attaching package: 'caret'
```

```
## The following object is masked from 'package:survival':
##
##     cluster
```

```r
set.seed(99)
cancer_set <- bcancer[, -1]
total_set <- createDataPartition(cancer_set$class, p = 0.60, list = FALSE)
training_set <- cancer_set[total_set,]
testing_set <- cancer_set[-total_set,]
```

Now we will run a tree model. A tree model breaks down all obervations from the training set by the key predictors which in this case are cell size and nuceli.

```r
library(rattle)
```

```
## Rattle: A free graphical interface for data science with R.
## Version 5.2.0 Copyright (c) 2006-2018 Togaware Pty Ltd.
## Type 'rattle()' to shake, rattle, and roll your data.
```

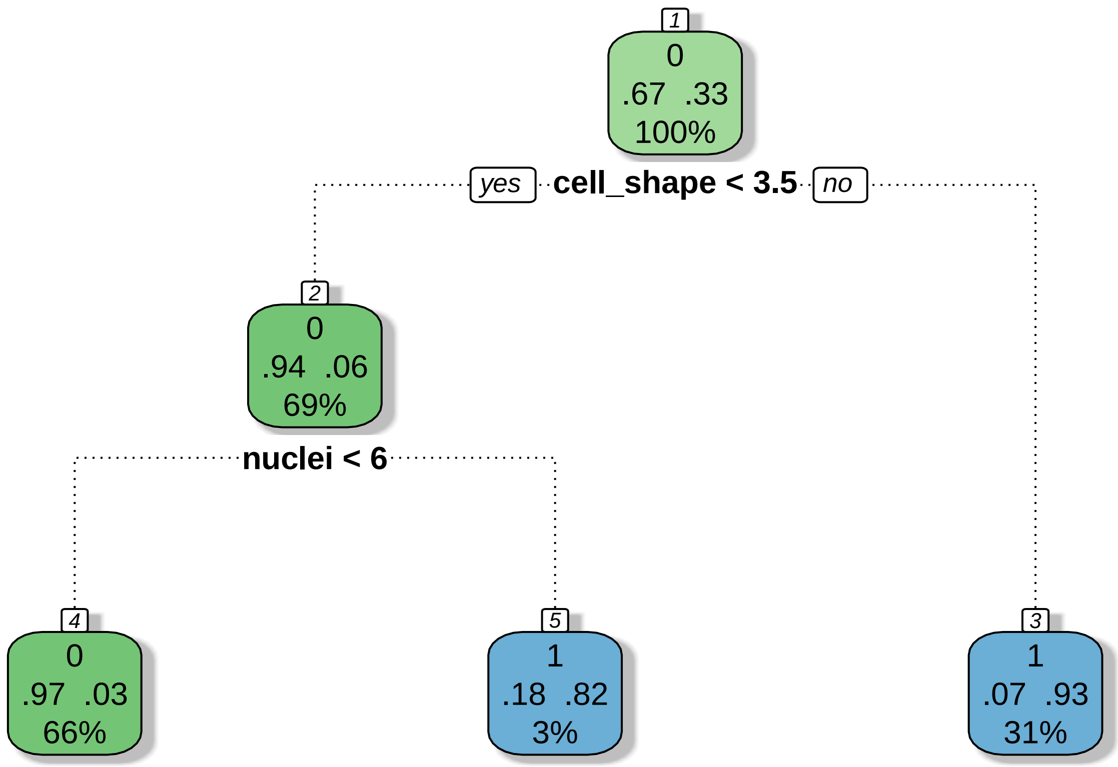
```r
attach(training_set)
```

```
## The following objects are masked from bcancer:
##
##     cell_shape, cell_size, chromatin, class, clump_thick,
##     epithelial, marginal, mitoses, nuclei, nucleoli
```

```r
training_set$class <- as.factor(training_set$class)
tree_model <- train(class~., data=training_set, method="rpart")
tree_model
```

```
## CART
##
## 420 samples
##   9 predictor
##   2 classes: '0', '1'
##
## No pre-processing
## Resampling: Bootstrapped (25 reps)
## Summary of sample sizes: 420, 420, 420, 420, 420, 420, ...
## Resampling results across tuning parameters:
##
##   cp          Accuracy   Kappa
##   0.00000000  0.9366472  0.8572399
##   0.05072464  0.9320596  0.8477359
##   0.81884058  0.8393058  0.5552945
##
## Accuracy was used to select the optimal model using the largest value.
## The final value used for the model was cp = 0.
```

```r
fancyRpartPlot(tree_model$finalModel)
```



Rattle 2019−Sep−11 12:09:35 muizrahemtullah

Now we will run a LDA Model using 10-Fold Cross Validation. The LDA algorithm uses a linear combinition of features, in this case the features of breast cancer like skin type, that characterizes the predictor variable.

```
control <- trainControl(method="cv", number=10)
metric <- "Accuracy"

set.seed(99)
lda_fit <- train(class~., data=training_set, method="lda", metric=metric, trControl=control)
lda_fit
```

```
## Linear Discriminant Analysis
##
## 420 samples
##   9 predictor
##   2 classes: '0', '1'
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 378, 377, 378, 379, 378, 378, ...
## Resampling results:
##
##   Accuracy   Kappa
##   0.9545823  0.895795
```

We will now use a KNN Machine Learning Algorithm. We attempt to predict the type of tumor by looking at the neighbors of the observation we are working on at that moment. In this context of the problem, this may not be the best model.

```r
set.seed(99)
knn_fit <- train(class~., data=training_set, method="knn", metric=metric, trControl=control)
knn_fit
```

```
## k-Nearest Neighbors
##
## 420 samples
##   9 predictor
##   2 classes: '0', '1'
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 378, 377, 378, 379, 378, 378, ...
## Resampling results across tuning parameters:
##
##   k  Accuracy   Kappa
##   5  0.9689261  0.9296265
##   7  0.9665451  0.9239771
##   9  0.9713070  0.9348862
##
## Accuracy was used to select the optimal model using the largest value.
## The final value used for the model was k = 9.
```

We will now use a Bayesian Generalized Linear Model

```r
set.seed(99)
bayes_fit <- train(class~., data=training_set, method="bayesglm", metric=metric, trControl=control)
bayes_fit
```

```
## Bayesian Generalized Linear Model
##
## 420 samples
##   9 predictor
##   2 classes: '0', '1'
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 378, 377, 378, 379, 378, 378, ...
## Resampling results:
##
##   Accuracy  Kappa
##   0.968868  0.928889
```

We will now use an SVM Machine Learning Model. This model creates a boundary called a hyperplane and examines the observations closest to the boundary to classify.

```r
set.seed(99)
svm_fit <- train(class~., data=training_set, method="svmRadial", metric=metric, trControl=control)
svm_fit
```

```
## Support Vector Machines with Radial Basis Function Kernel
##
## 420 samples
##   9 predictor
##   2 classes: '0', '1'
##
```

```
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 378, 377, 378, 379, 378, 378, ...
## Resampling results across tuning parameters:
##
##   C     Accuracy   Kappa
##   0.25  0.9306593  0.8530663
##   0.50  0.9355347  0.8617385
##   1.00  0.9546404  0.9011021
##
## Tuning parameter 'sigma' was held constant at a value of 0.793333
## Accuracy was used to select the optimal model using the largest value.
## The final values used for the model were sigma = 0.793333 and C = 1.
```

We will now use a Random Forest Model. We take a random number of predictors and create models with them. After several times of using random predictors and creating several models, we average them out to create one powerful model using elements from each of the weaker submodels.

```
set.seed(99)
random_forest_fit <- train(class~., data=training_set, method="rf", metric=metric, trControl=control)
random_forest_fit
```

```
## Random Forest
##
## 420 samples
##   9 predictor
##   2 classes: '0', '1'
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 378, 377, 378, 379, 378, 378, ...
## Resampling results across tuning parameters:
##
##   mtry  Accuracy   Kappa
##   2     0.9761270  0.9460765
##   5     0.9664870  0.9241409
##   9     0.9592861  0.9077390
##
## Accuracy was used to select the optimal model using the largest value.
## The final value used for the model was mtry = 2.
```

Finally, we will use a XGB Linear Boosting Model. This model is based on the concept of Boosting. This is a method for reducing our variance. We obtain new datasets by drawing existing observations from our existing dataset with replacement. This way we can train our model further and average out all predictions. In addition, we grow our models sequentially and fit a model to the residuals of the previous model for further accuracy.

```
set.seed(99)
boosting_fit <- train(class~ ., data=training_set, method="xgbLinear", metric=metric, trControl=control)
boosting_fit
```

```
## eXtreme Gradient Boosting
##
## 420 samples
##   9 predictor
##   2 classes: '0', '1'
```

```
##
## No pre-processing
## Resampling: Cross-Validated (10 fold)
## Summary of sample sizes: 378, 377, 378, 379, 378, 378, ...
## Resampling results across tuning parameters:
##
##   lambda  alpha  nrounds  Accuracy   Kappa
##   0e+00   0e+00   50      0.9616671  0.9139308
##   0e+00   0e+00  100      0.9616671  0.9139308
##   0e+00   0e+00  150      0.9616671  0.9139308
##   0e+00   1e-04   50      0.9616671  0.9139308
##   0e+00   1e-04  100      0.9616671  0.9139308
##   0e+00   1e-04  150      0.9592280  0.9079579
##   0e+00   1e-01   50      0.9592280  0.9079579
##   0e+00   1e-01  100      0.9592280  0.9079579
##   0e+00   1e-01  150      0.9592280  0.9079579
##   1e-04   0e+00   50      0.9616671  0.9139308
##   1e-04   0e+00  100      0.9616671  0.9139308
##   1e-04   0e+00  150      0.9616671  0.9139308
##   1e-04   1e-04   50      0.9616671  0.9139308
##   1e-04   1e-04  100      0.9616671  0.9139308
##   1e-04   1e-04  150      0.9592280  0.9079579
##   1e-04   1e-01   50      0.9592280  0.9079579
##   1e-04   1e-01  100      0.9592280  0.9079579
##   1e-04   1e-01  150      0.9592280  0.9079579
##   1e-01   0e+00   50      0.9640480  0.9193854
##   1e-01   0e+00  100      0.9640480  0.9193854
##   1e-01   0e+00  150      0.9640480  0.9193854
##   1e-01   1e-04   50      0.9640480  0.9193854
##   1e-01   1e-04  100      0.9640480  0.9193854
##   1e-01   1e-04  150      0.9640480  0.9193854
##   1e-01   1e-01   50      0.9616671  0.9139308
##   1e-01   1e-01  100      0.9616671  0.9139308
##   1e-01   1e-01  150      0.9616671  0.9139308
##
## Tuning parameter 'eta' was held constant at a value of 0.3
## Accuracy was used to select the optimal model using the largest value.
## The final values used for the model were nrounds = 50, lambda = 0.1,
##  alpha = 0 and eta = 0.3.
```

The best model is the one with the highest kappa value. Recall that kappa represents the accuracy of the model, which is based on the Observed Accuracy and the Expected Accuracy. The Observed Accuracy is defined to be all instances where the machine learning model was in agreement with the ground truth. In the context of this problem, the Observed Accuracy is all instances where the model's prediction of weather the person has a benign or malignant tumor was in agreement of what type of tumor the patient had in real life. It is the number of times the model was correct. The Expected Accuracy is based on the number of times we classify the type of cancer according to ground truth multiplied by the number of times we classify the type of cancer according to ground truth. Then kappa is equal to $\kappa = \frac{O-E}{1-E}$, where O is Oberved Accuracy and E is Expected Accuracy. The kappa values for each model are as follows:

- The kappa value for the XGB Linear Boosting Model is .912
- The kappa value for the Random Forest Model is .923
- The kappa value for the SVM Model is .873

- The kappa value for the Bayes Model is .895
- The kappa value for the KNN Model is .159
- The kappa value for the LDA Model is .905

Judging by these values, it seems that the Random Forest Model is the best model followed closely by the XGB Linear Boosting Model and the LDA Model.