

BMEG 310 Assignment 2

Wesley Chan, Flora Deng, Mary Graves

2022-10-17

```
# libraries
library(ggplot2)
library(ROCR)
```

```
# loading in data
ovarian.dataset <- read.delim("ovarian.data", sep=",", header = FALSE)
features <- c("perimeter", "area", "smoothness", "symmetry", "concavity",
paste("protein", seq(1, 25) ))
names(ovarian.dataset) <- c("cell_id", "diagnosis", features)
```

```
# check data (REMOVE FOR SUBMISSION)
head(ovarian.dataset)
```

```
##      cell_id diagnosis perimeter   area smoothness symmetry concavity protein 1
## 1    842302         M    17.99 10.38    122.80   1001.0    0.11840    0.27760
## 2    842517         M    20.57 17.77    132.90   1326.0    0.08474    0.07864
## 3   84300903         M    19.69 21.25    130.00   1203.0    0.10960    0.15990
## 4   84348301         M    11.42 20.38     77.58    386.1    0.14250    0.28390
## 5   84358402         M    20.29 14.34    135.10   1297.0    0.10030    0.13280
## 6    843786         M    12.45 15.70     82.57    477.1    0.12780    0.17000
##      protein 2 protein 3 protein 4 protein 5 protein 6 protein 7 protein 8
## 1      0.3001   0.14710   0.2419   0.07871    1.0950    0.9053    8.589
## 2      0.0869   0.07017   0.1812   0.05667    0.5435    0.7339    3.398
## 3      0.1974   0.12790   0.2069   0.05999    0.7456    0.7869    4.585
## 4      0.2414   0.10520   0.2597   0.09744    0.4956    1.1560    3.445
## 5      0.1980   0.10430   0.1809   0.05883    0.7572    0.7813    5.438
## 6      0.1578   0.08089   0.2087   0.07613    0.3345    0.8902    2.217
##      protein 9 protein 10 protein 11 protein 12 protein 13 protein 14 protein 15
## 1     153.40   0.006399   0.04904   0.05373    0.01587    0.03003    0.006193
## 2      74.08   0.005225   0.01308   0.01860    0.01340    0.01389    0.003532
## 3      94.03   0.006150   0.04006   0.03832    0.02058    0.02250    0.004571
## 4      27.23   0.009110   0.07458   0.05661    0.01867    0.05963    0.009208
## 5      94.44   0.011490   0.02461   0.05688    0.01885    0.01756    0.005115
## 6      27.19   0.007510   0.03345   0.03672    0.01137    0.02165    0.005082
##      protein 16 protein 17 protein 18 protein 19 protein 20 protein 21 protein 22
## 1      25.38    17.33    184.60    2019.0    0.1622    0.6656    0.7119
## 2      24.99    23.41    158.80    1956.0    0.1238    0.1866    0.2416
## 3      23.57    25.53    152.50    1709.0    0.1444    0.4245    0.4504
## 4      14.91    26.50     98.87     567.7    0.2098    0.8663    0.6869
## 5      22.54    16.67    152.20    1575.0    0.1374    0.2050    0.4000
## 6      15.47    23.75    103.40     741.6    0.1791    0.5249    0.5355
```

```
##   protein 23 protein 24 protein 25
## 1    0.2654    0.4601    0.11890
## 2    0.1860    0.2750    0.08902
## 3    0.2430    0.3613    0.08758
## 4    0.2575    0.6638    0.17300
## 5    0.1625    0.2364    0.07678
## 6    0.1741    0.3985    0.12440
```

```
# Question 1 - Dimensionality Reduction
```

```
# Question 2 - Clustering
```

```
# Question 3 - Classification
```

```
# Divide data into training set and test set
```

```
ovarian.dataset.train <- ovarian.dataset[sample(nrow(ovarian.dataset))
[1:(nrow(ovarian.dataset)/2)],]
ovarian.dataset.test <- ovarian.dataset[sample(nrow(ovarian.dataset))
[(nrow(ovarian.dataset)/2):(nrow(ovarian.dataset))],]
```

```
# Contribution Statement
```

```
# Wesley Chan (75520023): did literally nothing
```

```
# Flora Deng ():
```

```
# Mary Graves ():
```