



Naive Bayes

Breast Cancer. Use the 'wisc_bc_data.csv' dataset and build a Naïve Bayes model to predict 'diagnosis' cases M: malignant or B: benign depending on all the numerical predictors. Use `set.seed(1234)` and 455 cases to train the model. Predict cases of malignant and benign in test dataset. Calculate the confusion matrix

```
> wbcd = read.csv(file.choose(), header = TRUE)
> str(wbcd)
'data.frame': 569 obs. of 32 variables:
 $ id      : int  842302 842517 84300903 84348301 84358402 843786 844359 84458
202 844981 84501001 ...
 $ diagnosis : chr  "M" "M" "M" "M" ...
 $ radius_mean : num  18 20.6 19.7 11.4 20.3 ...
 $ texture_mean : num  10.4 17.8 21.2 20.4 14.3 ...
 $ perimeter_mean : num  122.8 132.9 130 77.6 135.1 ...
 $ area_mean : num  1001 1326 1203 386 1297 ...
 $ smoothness_mean : num  0.1184 0.0847 0.1096 0.1425 0.1003 ...
 $ compactness_mean : num  0.2776 0.0786 0.1599 0.2839 0.1328 ...
 $ concavity_mean : num  0.3001 0.0869 0.1974 0.2414 0.198 ...
 $ concave.points_mean : num  0.1471 0.0702 0.1279 0.1052 0.1043 ...
 $ symmetry_mean : num  0.242 0.181 0.207 0.26 0.181 ...
 $ fractal_dimension_mean : num  0.0787 0.0567 0.06 0.0974 0.0588 ...
 $ radius_se : num  1.095 0.543 0.746 0.496 0.757 ...
 $ texture_se : num  0.905 0.734 0.787 1.156 0.781 ...
 $ perimeter_se : num  8.59 3.4 4.58 3.44 5.44 ...
 $ area_se : num  153.4 74.1 94 27.2 94.4 ...
 $ smoothness_se : num  0.0064 0.00522 0.00615 0.00911 0.01149 ...
 $ compactness_se : num  0.049 0.0131 0.0401 0.0746 0.0246 ...
 $ concavity_se : num  0.0537 0.0186 0.0383 0.0566 0.0569 ...
 $ concave.points_se : num  0.0159 0.0134 0.0206 0.0187 0.0188 ...
 $ symmetry_se : num  0.03 0.0139 0.0225 0.0596 0.0176 ...
 $ fractal_dimension_se : num  0.00619 0.00353 0.00457 0.00921 0.00511 ...
 $ radius_worst : num  25.4 25 23.6 14.9 22.5 ...
 $ texture_worst : num  17.3 23.4 25.5 26.5 16.7 ...
 $ perimeter_worst : num  184.6 158.8 152.5 98.9 152.2 ...
 $ area_worst : num  2019 1956 1709 568 1575 ...
 $ smoothness_worst : num  0.162 0.124 0.144 0.21 0.137 ...
 $ compactness_worst : num  0.666 0.187 0.424 0.866 0.205 ...
 $ concavity_worst : num  0.712 0.242 0.45 0.687 0.4 ...
 $ concave.points_worst : num  0.265 0.186 0.243 0.258 0.163 ...
 $ symmetry_worst : num  0.46 0.275 0.361 0.664 0.236 ...
 $ fractal_dimension_worst : num  0.1189 0.089 0.0876 0.173 0.0768 ...
> wbcd = wbcd[, -1]
> set.seed(1234)
> s = sample(569, 455)
> train = wbcd[s,]
> test = wbcd[-s,]
> train$diagnosis = as.factor(train$diagnosis)
> test$diagnosis = as.factor(test$diagnosis)
> fit = naiveBayes(diagnosis ~ ., data = train)
```

```

> summary(fit)
      Length Class  Mode
apriori     2    table numeric
tables     30  -none- list
levels      2  -none- character
isnumeric  30  -none- logical
call        4  -none- call
> pred = predict(fit,test)
> table(Actual=test$diagnosis,predicted =pred)
      predicted
Actual  B   M
      B 69   6
      M  6  33

```