Practical - 4 Shivam Tawari A-58

Aim: Write a program in R for implementation of Bayes Theorem.

Theory:

Naive Bayes:
Naive Bayes is a Supervised Machine Learning Algorithm based on the Bayes Theorem that ~~is used~~ the predictor variables in Machine Learning model are independent of each other. Meaning that the outcome of a model depends on a set of independent varibles that have nothing to do with each other.

In real-world problems, predictor variables aren't always independent of each other, there are always some correlations between them. Since, Naive Bayes considers each predictor variable to be independent of any other variable in the model, it is called 'Naive'.

Principle:
The principle behind Naive Bayes is the Bayes theorem also known as the Bayes Rule. The Bayes theorem is used to calculate

the conditional probability, which is nothing but the probability of an event occuring based on information about the events in the past.

Mathematically, the Bayes Theorem is represented as:

$$P(A|B) = \frac{P(B/A)\,P(A)}{P(B)}$$

Derivation:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{---- eq ①}$$

Probability of A given B.

$$P(B/A) = \frac{P(B \cap A)}{P(A)} \quad \text{---- eq ②}$$

Probability of B given A.

$$\boxed{P(A|B) = \frac{P(B|A)\;P(A)}{P(B)}}$$

Bayes Thm. - Naive Bayes.

Consider a dataset 1500 obs.

Output classes :

Cat

Parrot

Turtle

Predictor variables :

Swim

Wings

Green Colour

Sharp teeth

Type	Swim	Wings	Green	Sharp teeth
Cat	450/500	0	0	500/500
Parrot	50/500	500/500	400/500	0
Turtle	500/1500	0	100/500	500/500

Code :

```
install.packages ("e1071")
install.packages ('ggplot2')
install.packages ('caret')
library ("caret")
library ("ggplot2")
library ("e1071")
```

```
head (iris)
names(iris)

x = iris[,-5]
y = iris$Species

model
model = train(x,y, 'nb', trControl = train(control
        (method = 'cv', number = 10))
predict (model$$)
table (predict (model$finalModel, x)$class, y)
```

Conclusion: Hence, successfully implemented the Bayes Theorem in R program.

Code:

Practical 4.R

```r
#Shivam Tawari A-58
install.packages('e1071')
install.packages('ggplot2')
install.packages('caret')
library("caret")
library("ggplot2")
library("e1071")

head(iris)
names(iris)

x = iris[,-5]
y = iris$Species

model
model = train(x,y,'nb',trControl = trainControl(method = 'cv',number = 10))
predict(model$f)

table(predict(model$finalModel,x)$class,y)

```

Output:

```
> head(iris)
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1         3.5          1.4         0.2  setosa
2          4.9         3.0          1.4         0.2  setosa
3          4.7         3.2          1.3         0.2  setosa
4          4.6         3.1          1.5         0.2  setosa
5          5.0         3.6          1.4         0.2  setosa
6          5.4         3.9          1.7         0.4  setosa
> library("e1071")
> head(iris)
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1         3.5          1.4         0.2  setosa
2          4.9         3.0          1.4         0.2  setosa
3          4.7         3.2          1.3         0.2  setosa
4          4.6         3.1          1.5         0.2  setosa
5          5.0         3.6          1.4         0.2  setosa
6          5.4         3.9          1.7         0.4  setosa
> names(iris)
[1] "Sepal.Length" "Sepal.Width"  "Petal.Length" "Petal.Width"
[5] "Species"
> x = iris[,-5]
> y = iris$Species
> model
Naive Bayes

150 samples
  4 predictor
  3 classes: 'setosa', 'versicolor', 'virginica'
```

```
Naive Bayes

150 samples
  4 predictor
  3 classes: 'setosa', 'versicolor', 'virginica'

No pre-processing
Resampling: Cross-Validated (10 fold)
Summary of sample sizes: 135, 135, 135, 135, 135, 135, ...
Resampling results across tuning parameters:

  usekernel  Accuracy  Kappa
  FALSE      0.96      0.94
   TRUE      0.96      0.94

Tuning parameter 'fL' was held constant at a value of 0
Tuning
 parameter 'adjust' was held constant at a value of 1
Accuracy was used to select the optimal model using the largest value.
The final values used for the model were fL = 0, usekernel = FALSE
 and adjust = 1.
> model = train(x,y,'nb',trControl = trainControl(method = 'cv',number = 10))
> predict(model$f)
$class
        X1         X2         X3         X4         X5         X6
    setosa     setosa     setosa     setosa     setosa     setosa
        X7         X8         X9        X10        X11        X12
    setosa     setosa     setosa     setosa     setosa     setosa
       X13        X14        X15        X16        X17        X18
    setosa     setosa     setosa     setosa     setosa     setosa
/cloud/project/
      X109       X110       X111       X112       X113       X114
 virginica  virginica  virginica  virginica  virginica  virginica
      X115       X116       X117       X118       X119       X120
 virginica  virginica  virginica  virginica  virginica versicolor
      X121       X122       X123       X124       X125       X126
 virginica  virginica  virginica  virginica  virginica  virginica
      X127       X128       X129       X130       X131       X132
 virginica  virginica  virginica  virginica  virginica  virginica
      X133       X134       X135       X136       X137       X138
 virginica versicolor  virginica  virginica  virginica  virginica
      X139       X140       X141       X142       X143       X144
 virginica  virginica  virginica  virginica  virginica  virginica
      X145       X146       X147       X148       X149       X150
 virginica  virginica  virginica  virginica  virginica  virginica
Levels: setosa versicolor virginica

$posterior
           setosa   versicolor    virginica
X1   1.000000e+00 3.122328e-09 8.989129e-11
X2   9.999999e-01 4.953302e-08 1.361560e-09
X3   1.000000e+00 1.949717e-08 1.152761e-09
X4   1.000000e+00 1.146273e-08 6.616756e-10
X5   1.000000e+00 8.839954e-10 8.567477e-11
X6   1.000000e+00 3.818715e-09 5.965843e-09
X7   1.000000e+00 7.394006e-09 6.702907e-10
X8   1.000000e+00 5.311568e-09 1.920277e-10
X9   1.000000e+00 6.502476e-09 3.193962e-10
X10  9.999998e-01 1.731985e-07 5.531788e-09
X11  1.000000e+00 1.233528e-09 4.372981e-10
X12  1.000000e+00 6.936685e-09 4.552987e-10
```

/cloud/project/

```
X129 3.101469e-09 1.442934e-06 9.999986e-01
X130 6.122836e-08 3.400634e-04 9.996599e-01
X131 1.389474e-08 2.976597e-06 9.999970e-01
X132 8.766738e-07 2.312170e-10 9.999991e-01
X133 3.229494e-09 1.502497e-06 9.999985e-01
X134 5.830326e-09 5.463686e-01 4.536314e-01
X135 2.147851e-08 2.848673e-02 9.715132e-01
X136 4.877334e-08 3.680347e-09 9.999999e-01
X137 6.367711e-08 9.930942e-07 9.999989e-01
X138 1.041818e-08 5.448493e-04 9.994551e-01
X139 5.004383e-07 2.024076e-01 7.975919e-01
X140 1.996159e-08 1.165611e-05 9.999883e-01
X141 1.999575e-08 1.295204e-06 9.999987e-01
X142 2.161884e-08 1.231467e-04 9.998768e-01
X143 4.982326e-07 2.784325e-02 9.721562e-01
X144 3.349443e-08 4.857252e-07 9.999995e-01
X145 7.668677e-08 7.172953e-07 9.999992e-01
X146 1.158419e-08 8.914668e-05 9.999108e-01
X147 7.841053e-10 2.133359e-02 9.786664e-01
X148 7.966669e-09 3.437503e-04 9.996562e-01
X149 5.761966e-08 1.351426e-05 9.999864e-01
X150 1.368634e-06 5.710392e-02 9.428947e-01

> table(predict(model$finalModel,x)$class,y)
            y
             setosa versicolor virginica
  setosa         50          0         0
  versicolor      0         47         3
  virginica       0          3        47
>
```