Experiment No:-7

Objective:- To apply Bayes’ Theorem for estimating the posterior probability and Principle Component Analysis.

Bayes’ Theory

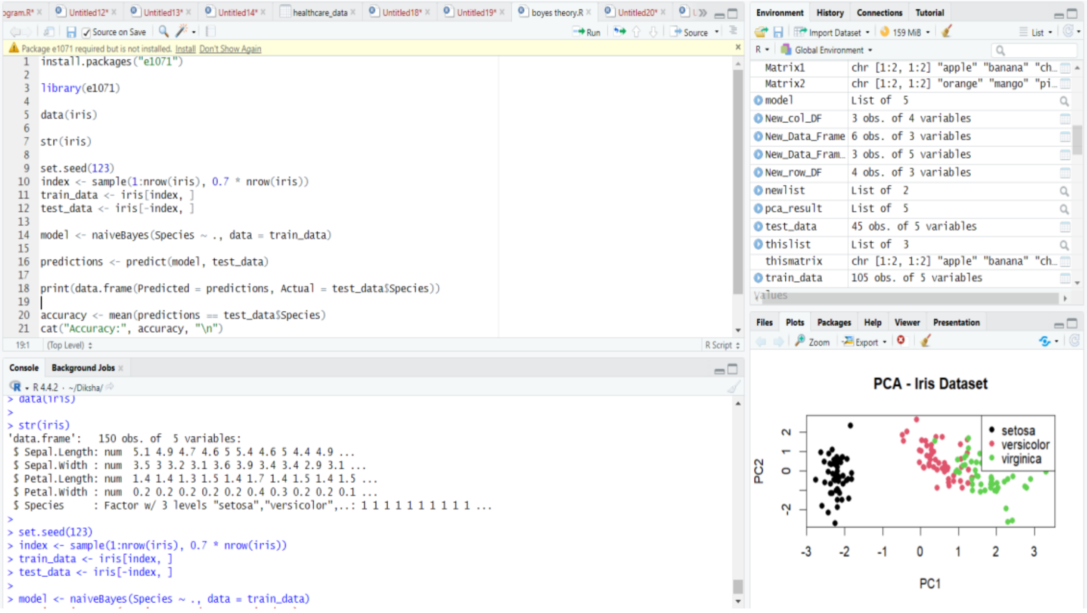
Bayes’ Theorem is a fundamental concept in probability theory and statistics that describes how to update the probability of a hypothesis based on new evidence. It allows us to calculate the posterior probability — the probability of an event occurring given that another event has already occurred.

The theorem is mathematically expressed as:

P(A|B)=P(B|A)·P(A)P(B)P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}P(A|B)=P(B)P(B|A)·P(A)

Where:

- P(A|B)P(A|B)P(A|B) = Posterior probability (probability of event A given event B)
- P(B|A)P(B|A)P(B|A) = Likelihood (probability of event B given event A)
- P(A)P(A)P(A) = Prior probability of event A
- P(B)P(B)P(B) = Total probability of event B



Package e1071 required but is not installed. Install Don't Show Again

```r
install.packages("e1071")

library(e1071)

data(iris)

str(iris)

set.seed(123)
index <- sample(1:nrow(iris), 0.7 * nrow(iris))
train_data <- iris[index, ]
test_data <- iris[-index, ]

model <- naiveBayes(Species ~ ., data = train_data)

predictions <- predict(model, test_data)

print(data.frame(Predicted = predictions, Actual = test_data$Species))

accuracy <- mean(predictions == test_data$Species)
cat("Accuracy:", accuracy, "\n")
```

```
> data(iris)
>
> str(iris)
'data.frame':   150 obs. of  5 variables:
 $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
 $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
 $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
 $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
 $ Species     : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...
>
> set.seed(123)
> index <- sample(1:nrow(iris), 0.7 * nrow(iris))
> train_data <- iris[index, ]
> test_data <- iris[-index, ]
>
> model <- naiveBayes(Species ~ ., data = train_data)
```

Matrix1	chr [1:2, 1:2] "apple" "banana" "ch…
Matrix2	chr [1:2, 1:2] "orange" "mango" "pi…
model	List of 5
New_col_DF	3 obs. of 4 variables
New_Data_Frame	6 obs. of 3 variables
New_Data_Fram…	3 obs. of 5 variables
New_row_DF	4 obs. of 3 variables
newlist	List of 2
pca_result	List of 5
test_data	45 obs. of 5 variables
thislist	List of 3
thismatrix	chr [1:2, 1:2] "apple" "banana" "ch…
train_data	105 obs. of 5 variables

Principle Component Analysis (PCA)

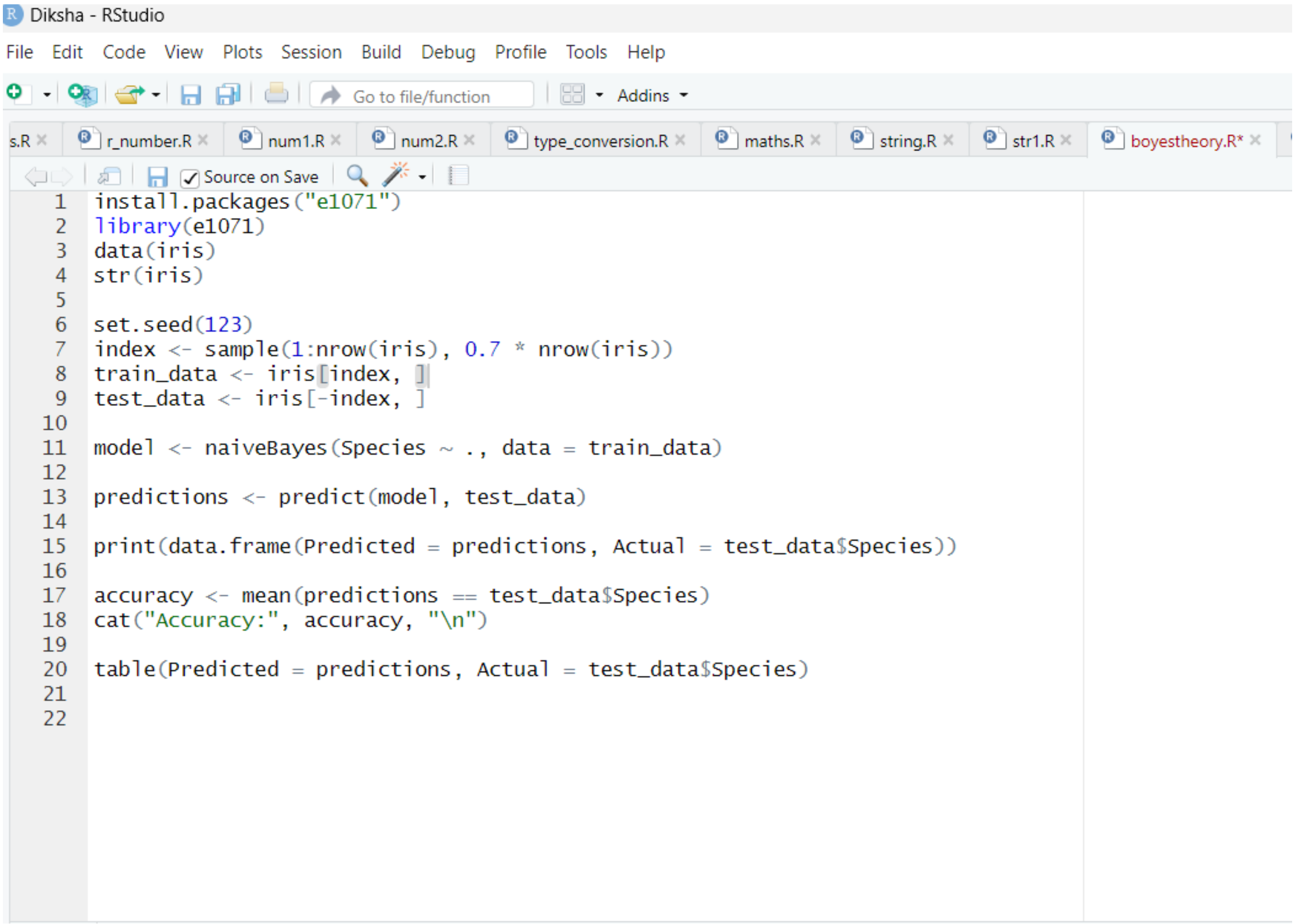
Principal Component Analysis (PCA) is a statistical technique used for dimensionality reduction while preserving as much variance in the data as possible. It transforms the original set of correlated variables into a new set of uncorrelated variables called principal components. These components are ordered such that the first few retain most of the variation present in the original variables.

The key idea is to project high-dimensional data onto a lower-dimensional subspace to:

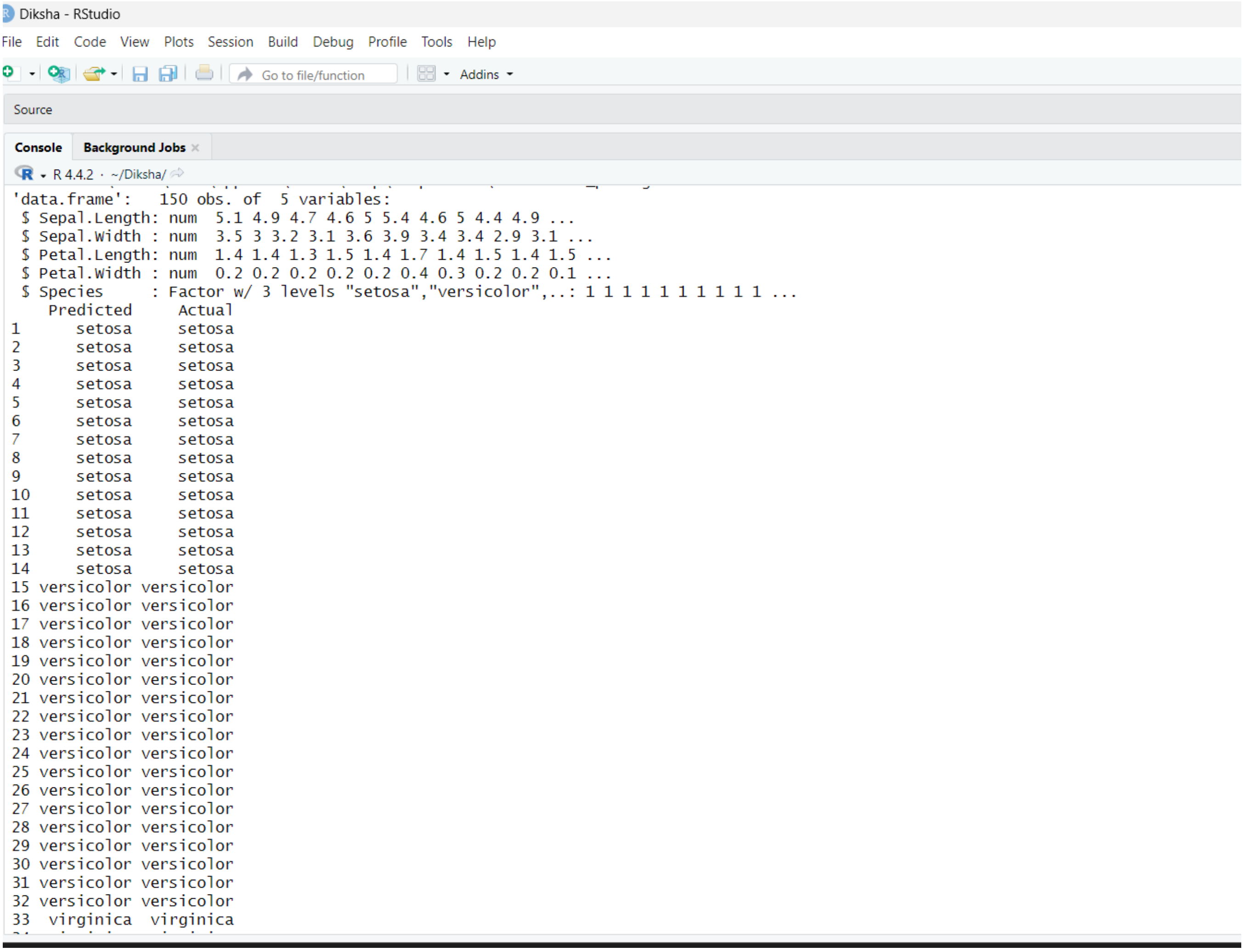
- Simplify analysis and visualization
- Reduce noise and redundancy
- Improve computational efficiency for machine learning tasks

PCA achieves this by:

1. Standardizing the data (if necessary)
2. Computing the covariance matrix
3. Extracting the eigenvalues and eigenvectors
4. Selecting top k principal components based on variance
5. Transforming the data into the new subspace



```r
install.packages("e1071")
library(e1071)
data(iris)
str(iris)

set.seed(123)
index <- sample(1:nrow(iris), 0.7 * nrow(iris))
train_data <- iris[index, ]
test_data <- iris[-index, ]

model <- naiveBayes(Species ~ ., data = train_data)

predictions <- predict(model, test_data)

print(data.frame(Predicted = predictions, Actual = test_data$Species))

accuracy <- mean(predictions == test_data$Species)
cat("Accuracy:", accuracy, "\n")

table(Predicted = predictions, Actual = test_data$Species)
```



```
'data.frame':	150 obs. of  5 variables:
 $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
 $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
 $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
 $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
 $ Species     : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...
    Predicted     Actual
1      setosa     setosa
2      setosa     setosa
3      setosa     setosa
4      setosa     setosa
5      setosa     setosa
6      setosa     setosa
7      setosa     setosa
8      setosa     setosa
9      setosa     setosa
10     setosa     setosa
11     setosa     setosa
12     setosa     setosa
13     setosa     setosa
14     setosa     setosa
15 versicolor versicolor
16 versicolor versicolor
17 versicolor versicolor
18 versicolor versicolor
19 versicolor versicolor
20 versicolor versicolor
21 versicolor versicolor
22 versicolor versicolor
23 versicolor versicolor
24 versicolor versicolor
25 versicolor versicolor
26 versicolor versicolor
27 versicolor versicolor
28 versicolor versicolor
29 versicolor versicolor
30 versicolor versicolor
31 versicolor versicolor
32 versicolor versicolor
33  virginica  virginica
```