43260_PCA.r

tanmay

2021-11-23

```
getwd()
```

```
## [1] "/home/tanmay/Downloads/CL-VII/MLA/Assignment6"
```

```
df <- read.csv("iris.csv", header=TRUE, sep=",")
head(df)
```

```
##   X Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1 1          5.1         3.5          1.4         0.2  setosa
## 2 2          4.9         3.0          1.4         0.2  setosa
## 3 3          4.7         3.2          1.3         0.2  setosa
## 4 4          4.6         3.1          1.5         0.2  setosa
## 5 5          5.0         3.6          1.4         0.2  setosa
## 6 6          5.4         3.9          1.7         0.4  setosa
```

```
x <- df[c("Sepal.Length", "Sepal.Width", "Petal.Length", "Petal.Width")]
head(x)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width
## 1          5.1         3.5          1.4         0.2
## 2          4.9         3.0          1.4         0.2
## 3          4.7         3.2          1.3         0.2
## 4          4.6         3.1          1.5         0.2
## 5          5.0         3.6          1.4         0.2
## 6          5.4         3.9          1.7         0.4
```

```
y <- df[c("Species")]
head(y)
```

```
##   Species
## 1  setosa
## 2  setosa
## 3  setosa
## 4  setosa
## 5  setosa
## 6  setosa
```

```
df.pr <- prcomp(x, center = TRUE, scale = TRUE)
summary(df.pr)
```

```
## Importance of components:
##                           PC1    PC2     PC3     PC4
## Standard deviation     1.7084 0.9560 0.38309 0.14393
## Proportion of Variance 0.7296 0.2285 0.03669 0.00518
## Cumulative Proportion  0.7296 0.9581 0.99482 1.00000
```

```
var_explained <- df.pr$sdev^2 / sum(df.pr$sdev^2)
print(var_explained)
```

```
## [1] 0.729624454 0.228507618 0.036689219 0.005178709
```

```
screeplot(df.pr, type = "l", npcs = 4, main = "Screeplot of the PCs")
abline(h = 1, col="red", lty=5)
legend("topright", legend=c("Eigenvalue = 1"), col=c("red"), lty=5, cex=0.6)
```