PCA

Bibek Sapkota

PCA

Loading the dataset and printing its first 6th values

```
df <- iris[, -5]
head(df)
```

```
##   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
```

Center and scale the data

```
df.scaled <- scale(df, center = TRUE, scale = TRUE)
```

Compute the correlation matrix:

```
# 1. Correlation matrix
res.cor <- cor(df.scaled)
round(res.cor, 2)
```

```
##              Sepal.Length Sepal.Width Petal.Length Petal.Width
## Sepal.Length         1.00       -0.12         0.87        0.82
## Sepal.Width         -0.12        1.00        -0.43       -0.37
## Petal.Length         0.87       -0.43         1.00        0.96
## Petal.Width          0.82       -0.37         0.96        1.00
```

Calculate the eigenvectors/eigenvalues of the correlation matrix:

```
# 2. Calculate eigenvectors/eigenvalues
res.eig <- eigen(res.cor)
res.eig
```

```
## eigen() decomposition
## $values
## [1] 2.91849782 0.91403047 0.14675688 0.02071484
##
## $vectors
##            [,1]        [,2]       [,3]       [,4]
## [1,]  0.5210659 -0.37741762  0.7195664  0.2612863
## [2,] -0.2693474 -0.92329566 -0.2443818 -0.1235096
## [3,]  0.5804131 -0.02449161 -0.1421264 -0.8014492
## [4,]  0.5648565 -0.06694199 -0.6342727  0.5235971
```

Compute projection matrix

task 1:Define the dimensions of new data set

```
k <- 3
```

task 2:Compute projection matrix (W)

```
W <- res.eig$vectors[,1:k]
```

compute the new dataset:

task 1: Using matrix calculation

```
df.new = df.scaled %*% W
```

task 2:Assign new column names

```
colnames(df.new) <- c("PC1", "PC2", "PC3")
head(df.new)
```

```
##            PC1        PC2         PC3
## [1,] -2.257141 -0.4784238  0.12727962
## [2,] -2.074013  0.6718827  0.23382552
## [3,] -2.356335  0.3407664 -0.04405390
## [4,] -2.291707  0.5953999 -0.09098530
## [5,] -2.381863 -0.6446757 -0.01568565
## [6,] -2.068701 -1.4842053 -0.02687825
```

Alternative method to compute new dataset

task 1: Transpose the intial data

```
df.scaled.t <- t(df.scaled)
```

task 2: Transpose the projection matrix

```
W.t <- t(W)
```

task 3: Compute transpose of new dataset

```
df.new.t <- W.t %*% df.scaled.t
```

task 4: Compute the new dataset

```
df.new <- t(df.new.t)
```

task 5: Assign new column names

```
colnames(df.new) <- c("PC1", "PC2", "PC3")
head(df.new)
```

```
##             PC1        PC2         PC3
## [1,] -2.257141 -0.4784238  0.12727962
## [2,] -2.074013  0.6718827  0.23382552
## [3,] -2.356335  0.3407664 -0.04405390
## [4,] -2.291707  0.5953999 -0.09098530
## [5,] -2.381863 -0.6446757 -0.01568565
## [6,] -2.068701 -1.4842053 -0.02687825
```