MSA 2017

Data Analysis in R
Examples with Linear Algebra

Author:
Shaina Race

PCA VIA SINGULAR VALUE DECOMPOSITION

Observations on the Rows

Let's demonstrate the fact that PCA and SVD are equivalent by using a toy matrix example.

$$\mathbf{X} = \begin{array}{c} \\ obs_1 \\ obs_2 \\ obs_3 \end{array} \begin{array}{c} \begin{array}{ccc} v_1 & v_2 & v_3 \end{array} \\ \begin{pmatrix} 2 & 1 & 7 \\ 5 & 6 & 4 \\ 4 & 3 & 9 \end{pmatrix} \end{array}$$

Covariance PCA via SVD To run a covariance PCA using the SVD, we need only center the data.

```
> X=matrix(c(2,1,7,1,5,6,4,3,9),nrow=3,byrow=T)
> X.PCA=prcomp(X)
> X.centered = scale(X, center=T,scale=F)
> X.centered.SVD=svd(X.centered)
> u=X.centered.SVD$u
> d=diag(X.centered.SVD$d)
> v=X.centered.SVD$v
> (loadingsPCA = X.PCA$rotation)
```

```
           PC1        PC2           PC3
[1,]  0.5543951 -0.4389146 -7.071068e-01
[2,] -0.6207190 -0.7840331  2.220446e-16
[3,]  0.5543951 -0.4389146  7.071068e-01
```

```
> (loadingsSVD=v)
```

```
           [,1]       [,2]          [,3]
[1,]  0.5543951 -0.4389146 -7.071068e-01
[2,] -0.6207190 -0.7840331  2.220446e-16
[3,]  0.5543951 -0.4389146  7.071068e-01
```

```
> (scoresPCA = X.PCA$x)
```

```
            PC1        PC2           PC3
[1,]  0.8718414  1.8606759 -8.326673e-17
[2,] -2.7198250 -0.3976271  8.881784e-16
[3,]  1.8479837 -1.4630488  0.000000e+00
```

```
> (scoresSVD=u%*%d)
```

```
           [,1]       [,2]         [,3]
[1,]  0.8718414  1.8606759 2.142511e-16
[2,] -2.7198250 -0.3976271 2.142511e-16
[3,]  1.8479837 -1.4630488 2.142511e-16
```

Correlation PCA To run a correlation PCA using the SVD, we need only standardize the data. The only change in the code is the addition of the option scale=T in both the prcomp() and the scale() function

```
> X=matrix(c(2,1,7,1,5,6,4,3,9),nrow=3,byrow=T)
> X.PCA=prcomp(X, scale=T)
> X.stnd = scale(X, center=T,scale=T)
> X.stnd.SVD=svd(X.stnd)
> u=X.stnd.SVD$u
> d=diag(X.stnd.SVD$d)
> v=X.stnd.SVD$v
> (loadingsPCA = X.PCA$rotation)
```

```
            PC1        PC2           PC3
[1,]  0.6583691 -0.2579732 -7.071068e-01
[2,] -0.3648291 -0.9310745  1.110223e-16
[3,]  0.6583691 -0.2579732  7.071068e-01
```

```
> (loadingsSVD=v)
```

```
          [,1]       [,2]          [,3]
[1,]  0.6583691 -0.2579732 -7.071068e-01
[2,] -0.3648291 -0.9310745  1.110223e-16
[3,]  0.6583691 -0.2579732  7.071068e-01
```

```
> (scoresPCA = X.PCA$x)
```

```
            PC1        PC2          PC3
[1,]  0.07749332  1.0436632 8.326673e-17
[2,] -1.51417244 -0.4807196 2.220446e-16
[3,]  1.43667912 -0.5629436 3.330669e-16
```

```
> (scoresSVD=u%*%d)
```

```
            [,1]       [,2]          [,3]
[1,]  0.07749332  1.0436632 2.467473e-16
[2,] -1.51417244 -0.4807196 2.467473e-16
[3,]  1.43667912 -0.5629436 2.467473e-16
```