Birla Institute of Technology and Science, Pilani

BITS-F464: Machine Learning
2^{nd} Semester 2018-19

Labsheet-03: PCA

1 Principal Component Analysis

PCA is used to perform an orthogonal transformation that converts a set of observations having correlated attributes into a set of attributes of linearly uncorrelated variables called principal components.

1.1 Example

Use `mtcars` dataset, which is built into R. The dataset consists of data on 32 models of car. For each car, you have 11 features, as mpg (Fuel consumption), cyl (Number of cylinders), disp (Displacement), hp (Gross horsepower), drat (Rear axle ratio), wt (Weight), qsec (speed and acceleration), vs (Engine block), am (Transmission automatic or manual) gear (Number of forward gears), carb (Number of carburetors). Let us see what is there in `mtcars`

```
head(mtcars)
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

Two of the features vs and am are categorical so drop them.

```
d01 <- mtcars[,c(1:7,10,11)]
head(d01)
```

	mpg	cyl	disp	hp	drat	wt	qsec	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	3	2
Valiant	18.1	6	225	105	2.76	3.460	20.22	3	1

Relationship between every pair of attributes could be seen by `plot(d01)`. Let us apply PCA on the data and see the summary.

```
d01 <- mtcars[,c(1:7,10,11)]
d01.pca <- princomp(d01, cor=TRUE, score=TRUE)

summary(d01.pca)
Importance of components:
                          Comp.1       Comp.2       Comp.3
Standard deviation     134.3820274 37.54656204 3.0181295838
Proportion of Variance   0.9270116  0.07236743 0.0004676043
Cumulative Proportion    0.9270116  0.99937900 0.9998466012
                             Comp.4       Comp.5       Comp.6
Standard deviation     1.254845e+00 8.904901e-01 6.371404e-01
Proportion of Variance 8.083193e-05 4.070624e-05 2.083882e-05
Cumulative Proportion  9.999274e-01 9.999681e-01 9.999890e-01
                             Comp.7       Comp.8       Comp.9
Standard deviation     3.006062e-01 2.814188e-01 2.124807e-01
Proportion of Variance 4.638724e-06 4.065453e-06 2.317617e-06
Cumulative Proportion  9.999936e-01 9.999977e-01 1.000000e
```

It is interesting to note that the first component itself has 92.70% variance. Added with 2nd component it becomes. 99.93%. Therefore it looks like only two values could suffice to describe the data for most of the applications. Try to see plot of the components

```
plot (d01.pca)
```

There is a very important variable called *loading* that specifies how individual attributes contribute to the components.

```
d01.pca$loadings

Loadings:
     Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9
mpg                 0.984                0.133
cyl                       -0.237 -0.226  0.823 -0.405 -0.191  0.109
disp -0.900  0.435
hp   -0.435 -0.899
drat                             0.131 -0.238        -0.941  0.188
wt                         0.133  0.244  0.126  0.223  0.163  0.907
qsec                       0.911  0.207  0.203 -0.217 -0.104 -0.153
gear                      -0.130  0.273 -0.350 -0.845  0.201  0.171
carb               -0.104 -0.272  0.864  0.263  0.152        -0.277
```

Transformed components could be obtained using *scores*

```
d02 <- d01.pca$scores
head(d02)
                      Comp.1     Comp.2     Comp.3     Comp.4     Comp.5
Comp.6      Comp.7      Comp.8      Comp.9
Mazda RX4          79.596905  2.136219 -2.182082 -2.5746575  0.7113548
0.32047992  0.15750146 -0.07061906 -0.20970382
Mazda RX4 Wag      79.599050  2.151464 -2.243211 -2.0306999  0.8896159
0.46613109  0.09265244 -0.08697755 -0.06376904
Datsun 710        133.892150 -5.056248 -2.158275  0.3741768 -1.1584686 -1.0543914
0.05183307  0.14432772  0.12156318
Hornet 4 Drive     -8.517325 44.982954  1.241749  0.7295492 -0.4270038
0.09005525 -0.02287295  0.22005003 -0.22486520
```

```
Hornet Sportabout −128.685064 30.817359  3.347341 −0.5353913 −0.7143029
0.30387460 −0.13567971 −0.05864375 −0.21892521
Valiant             23.219555 35.103458 −3.250535  1.3477916 −0.8212154
0.16633269 −0.22661022  0.42519000 −0.08630430
```