linearAlgebra: Scaling and Distance

Ivan Jacob Agaloos Pesigan

Deviation - Mean Centering

The $n \times 1$ vector of deviations from the mean is given by

$$\boldsymbol{d} = \boldsymbol{x} - \bar{x} \tag{1}$$

where $\boldsymbol{x}$ is an $n \times 1$ column vector and $\bar{x}$ is the mean of $\boldsymbol{x}$.

```
d(x[, 1])
```

```
##        [,1]
##  [1,]  0.24
##  [2,]  0.04
##  [3,] -0.16
##  [4,] -0.26
##  [5,]  0.14
##  [6,]  0.54
##  [7,] -0.26
##  [8,]  0.14
##  [9,] -0.46
## [10,]  0.04
```

The $n \times k$ matrix of deviations from the mean is given by

$$D = X - \mathbf{1}_n \bar{\boldsymbol{x}} \tag{2}$$

where $\boldsymbol{X}$ is an $n \times 1$ matrix, $\mathbf{1}_n$ is an $n \times 1$ column vector of ones, and $\bar{\boldsymbol{x}}$ is the $k \times 1$ mean vector of $\boldsymbol{X}$.

```
d(x)
```

```
##    Sepal.Length Sepal.Width Petal.Length Petal.Width
## 1          0.24        0.19        -0.05       -0.02
## 2          0.04       -0.31        -0.05       -0.02
## 3         -0.16       -0.11        -0.15       -0.02
## 4         -0.26       -0.21         0.05       -0.02
## 5          0.14        0.29        -0.05       -0.02
## 6          0.54        0.59         0.25        0.18
## 7         -0.26        0.09        -0.05        0.08
## 8          0.14        0.09         0.05       -0.02
## 9         -0.46       -0.41        -0.05       -0.02
## 10         0.04       -0.21         0.05       -0.12
```

Standardized Score - Scaling

The standardized score of the vector $\boldsymbol{x}$ is given by

$$z = \frac{\boldsymbol{x} - \bar{x}}{s} \tag{3}$$

where $\bar{x}$ is the mean and s is the standard deviation of $\boldsymbol{x}$.

```r
z(x[, 1])
```

```
##             [,1]
##  [1,]  0.8237318
##  [2,]  0.1372886
##  [3,] -0.5491545
##  [4,] -0.8923761
##  [5,]  0.4805102
##  [6,]  1.8533965
##  [7,] -0.8923761
##  [8,]  0.4805102
##  [9,] -1.5788192
## [10,]  0.1372886
```

The standardized score for the j^{th} column of a matrix $\boldsymbol{X}$ is given by

$$z_j = \frac{\boldsymbol{x}_j - \bar{x}_j}{s_j} \tag{4}$$

where $\boldsymbol{x}_j$ is the j^{th} column of a matrix $\boldsymbol{X}$, $\bar{x}_j$ is the mean of $\boldsymbol{x}_j$, and s_j is the standard deviation of $\boldsymbol{x}_j$.

```r
z(x)
```

```
##    Sepal.Length Sepal.Width Petal.Length Petal.Width
## 1     0.8237318   0.6186158     -0.46291  -0.2535463
## 2     0.1372886  -1.0093205     -0.46291  -0.2535463
## 3    -0.5491545  -0.3581460     -1.38873  -0.2535463
## 4    -0.8923761  -0.6837333      0.46291  -0.2535463
## 5     0.4805102   0.9442031     -0.46291  -0.2535463
## 6     1.8533965   1.9209649      2.31455   2.2819165
```

```
## 7    -0.8923761   0.2930285     -0.46291   1.0141851
## 8     0.4805102   0.2930285      0.46291  -0.2535463
## 9    -1.5788192  -1.3349078     -0.46291  -0.2535463
## 10    0.1372886  -0.6837333      0.46291  -1.5212777
```

Δ^2 - Squared Mahalanobis Distance

The squared Mahalanobis distance of the vector $\boldsymbol{x}$ is given by

$$\Delta^2 = (\boldsymbol{x} - \boldsymbol{\mu})' \, \boldsymbol{\Sigma}^{-1} \, (\boldsymbol{x} - \boldsymbol{\mu}) \tag{5}$$

where $\boldsymbol{\mu}$ is the mean vector and $\boldsymbol{\Sigma}$ is the covariance matrix.

```
deltacapsq(
  x,
  mu  = colMeans(x),
  sigmacap = stats::cov(x)
)
```

```
##         1         2         3         4         5         6         7         8
## 2.1836824 6.0959842 2.3508914 2.6450391 3.3674652 7.2574119 3.7700132 0.7755954
##         9        10
## 3.0556619 4.4982552
```