MDS: Link Squared Distance Matrix and Gram Matrix in Practice

Lieven Clement

statOmics, Ghent University (https://statomics.github.io)

Contents

1 Data 1

2 Centering 1

3 Gram matrix 2

4 Squared Distance matrix 2

5 Link Gram Matrix and Squared Distance Matrix 3

6 Show that N cancels out when multiplying with H 3

1 Data

Part of the iris dataset

```
X <- iris[1:5,1:4] %>% as.matrix
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
```

2 Centering

$$\mathbf{H} = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^T,$$

$$\mathbf{X}_c = \mathbf{H}\mathbf{X}$$

$$\mathbf{H}\mathbf{X}_c = \mathbf{X}_c$$

```r
H <- diag(nrow(X)) - matrix(1/nrow(X),nrow=nrow(X),ncol=nrow(X))
Xc <- H%*%X
colMeans(Xc)
```

```
##  Sepal.Length   Sepal.Width  Petal.Length   Petal.Width
##  1.776357e-16 -8.881784e-17 -4.440892e-17  0.000000e+00
```

```r
H%*%Xc-Xc
```

```
##       Sepal.Length  Sepal.Width Petal.Length Petal.Width
## [1,] -1.665335e-16 8.326673e-17 4.440892e-17           0
## [2,] -1.804112e-16 1.110223e-16 4.440892e-17           0
## [3,] -1.665335e-16 8.326673e-17 4.163336e-17           0
## [4,] -1.665335e-16 8.326673e-17 4.163336e-17           0
## [5,] -1.665335e-16 1.110223e-16 4.440892e-17           0
```

3 Gram matrix

Gram matrix: $\mathbf{G} = \mathbf{X}\mathbf{X}^T$ Here we work on centered data so $\mathbf{G} = \mathbf{X}_c\mathbf{X}_c^T$

```r
G <- Xc%*%t(Xc)
```

4 Squared Distance matrix

$$\mathbf{D}_X = \mathbf{N} - 2\mathbf{X}\mathbf{X}^T + \mathbf{N}^T,$$

```r
N <- matrix(diag(G),nrow(Xc),nrow(Xc))
N
```

```
##       [,1]  [,2]  [,3]  [,4]  [,5]
## [1,] 0.106 0.106 0.106 0.106 0.106
## [2,] 0.080 0.080 0.080 0.080 0.080
## [3,] 0.042 0.042 0.042 0.042 0.042
## [4,] 0.110 0.110 0.110 0.110 0.110
## [5,] 0.122 0.122 0.122 0.122 0.122
```

```r
dist2 <- N-2*G+t(N)
dist2
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,] 0.00 0.29 0.26 0.42 0.02
## [2,] 0.29 0.00 0.09 0.11 0.37
## [3,] 0.26 0.09 0.00 0.06 0.26
## [4,] 0.42 0.11 0.06 0.00 0.42
## [5,] 0.02 0.37 0.26 0.42 0.00
```

```r
dist(Xc)^2
```

```
##      1    2    3    4
## 2 0.29
## 3 0.26 0.09
## 4 0.42 0.11 0.06
## 5 0.02 0.37 0.26 0.42
```

5 Link Gram Matrix and Squared Distance Matrix

$$\mathbf{G} = \mathbf{X}_c\mathbf{X}_c^T = -\frac{1}{2}\mathbf{H}\mathbf{D}_X\mathbf{H}$$

```r
G
```

```
##        [,1]   [,2]   [,3]   [,4]   [,5]
## [1,]  0.106 -0.052 -0.056 -0.102  0.104
## [2,] -0.052  0.080  0.016  0.040 -0.084
## [3,] -0.056  0.016  0.042  0.046 -0.048
## [4,] -0.102  0.040  0.046  0.110 -0.094
## [5,]  0.104 -0.084 -0.048 -0.094  0.122
```

```r
-1/2*H%*%dist2%*%H
```

```
##        [,1]   [,2]   [,3]   [,4]   [,5]
## [1,]  0.106 -0.052 -0.056 -0.102  0.104
## [2,] -0.052  0.080  0.016  0.040 -0.084
## [3,] -0.056  0.016  0.042  0.046 -0.048
## [4,] -0.102  0.040  0.046  0.110 -0.094
## [5,]  0.104 -0.084 -0.048 -0.094  0.122
```

```r
-1/2*H%*%dist2%*%H - G
```

```
##               [,1]          [,2]          [,3]         [,4]          [,5]
## [1,] -2.775558e-17 -7.632783e-17 -2.775558e-17 1.387779e-17  0.000000e+00
## [2,] -7.632783e-17 -5.551115e-17 -1.387779e-17 1.387779e-17 -2.775558e-17
## [3,] -2.081668e-17 -1.387779e-17  2.775558e-17 6.245005e-17  2.775558e-17
## [4,]  1.387779e-17  6.938894e-18  5.551115e-17 8.326673e-17  0.000000e+00
## [5,]  0.000000e+00 -4.163336e-17  2.081668e-17 0.000000e+00  4.163336e-17
```

6 Show that N cancels out when multiplying with H

$$\mathbf{H}\mathbf{N}\mathbf{H} = \mathbf{0}$$

$$\mathbf{H}\mathbf{N}^T\mathbf{H} = \mathbf{0}$$

```r
N%*%H
```

```
##               [,1]          [,2] [,3]         [,4] [,5]
## [1,]  0.000000e+00  0.000000e+00    0 0.000000e+00    0
## [2,] -6.938894e-18 -6.938894e-18    0 3.469447e-18    0
## [3,]  0.000000e+00  0.000000e+00    0 0.000000e+00    0
## [4,] -6.938894e-18 -6.938894e-18    0 6.938894e-18    0
## [5,] -6.938894e-18 -6.938894e-18    0 3.469447e-18    0
```

```r
H%*%t(N)
```

```
##      [,1]          [,2] [,3]          [,4]          [,5]
## [1,]    0 -6.938894e-18    0 -6.938894e-18 -6.938894e-18
## [2,]    0 -6.938894e-18    0 -6.938894e-18 -6.938894e-18
## [3,]    0  0.000000e+00    0  0.000000e+00  0.000000e+00
## [4,]    0  3.469447e-18    0  6.938894e-18  3.469447e-18
## [5,]    0  0.000000e+00    0  0.000000e+00  0.000000e+00
```