
El método del Principal Component Analysis

¿Como se calcula el método utilizando un lenguaje de programación? Si utilizamos matlab por ejemplo.

En Matlab para reservar memoria se crea una matriz.

Primero introducimos los datos

```
samples = [  
10.0, 10.7;  
10.4, 9.8;  
9.7, 10.0;  
9.7, 10.1;  
11.7, 11.5;  
11.0, 10.8;  
8.7, 8.8;  
9.5, 9.3;  
10.1, 9.4;  
9.6, 9.6;  
10.5, 10.4;  
9.2, 9.0;  
11.3, 11.6;  
10.1, 9.8;  
8.5, 9.2;  
  
];
```

Si utilizamos una sola instrucción

```
S = cov(samples);  
S
```

```
S =  
  
    0.7986    0.6793  
    0.6793    0.7343
```

Para hacerla en distintas instrucciones tenemos que reservar la memoria primero y después hacer la matriz. Algo parecido a lo siguiente

```
S = [ cov(samples(:,1),samples(:,1)) cov(samples(:,1),samples(:,2))  
      cov(samples(:,2),samples(:,1)) cov(samples(:,2),samples(:,2))];  
S
```

```
S =  
  
    0.7986    0.7986    0.7986    0.6793  
    0.7986    0.7986    0.6793    0.7343
```

0.7343	0.6793	0.7343	0.7343
0.6793	0.7986	0.7343	0.7343

Aunque aquí, tenemos que algunos términos se repiten, esto tiene que ver con la documentación de MatLab.

Para arreglar esto vemos lo siguiente

```
avgMatrix = repmat(mean(samples), size(samples,1), 1);
deviation = samples - avgMatrix;
S = (deviation' * deviation) / (size(samples, 1) - 1);
S
```

S =

0.7986	0.6793
0.6793	0.7343

Es mucho mas eficiente hacer ciclos que matrices, porque cuando se hacen matrices se gasta mucha memoria.

Implementamos el método

```
S = cov(samples);
```

Introducimos lo eigenvalores

```
l = [1.4465, 0.0864];
```

Operamos el primer eigenvector

```
t = (S - eye(2) * l(1));
a = -t(1,1)/t(1,2);
U = [1;a] / sqrt(1+a^2);
```

Calculamos el segundo eigenvector

```
t = (S - eye(2) * l(2));
a = -t(2,2)/t(2,1);
U = [1;a] / sqrt(1+a^2);
U
```

U =

0.7236
-0.6902

Lo que representan las columnas de U, es una matriz de rotación que nos dice cuanto rotaron los ejes originales. Cuando leemos las gráficas de PCA normalmente se genera una media (0,0), el centro esta en el origen del sistema cartesiano, es por esta razón que en la gráfica del libro (p 23) el sistema tiene como origen el valor de las medias (10,10).

Los elementos de la matriz U son los cosenos directores del sistema coordinado viejo con respecto al nuevo.

Published with MATLAB® R2021b