

PCA mintapélda Matlabban

A MATLAB matematikai programcsomag tartalmazza a főkomponens-analízis módszerét is. A `princomp(X)` függvény a könyvben megadott lépéseket követve előbb kiszámolja a kovarianciamátrixot, majd meghatározza a kapcsolódó sajátértékeket és sajátvektorokat. Az eljárás bemenetként olyan adatmátrixot vár, amelyben az adatok várható értéke minden koordináta mentén zérus. Az **X** mátrixban a vektorokat sorvektorként kell megadni. Példánkban az alábbi kétdimenziós vektorokból indulunk ki:

$$\langle 1.0, 1.8 \rangle, \langle 2.1, 4.2 \rangle, \langle 0.7, 1.6 \rangle, \langle 0.5, 1.1 \rangle, \\ \langle 1.3, 2.5 \rangle, \langle 2.3, 4.4 \rangle, \langle 1.8, 3.5 \rangle, \langle 1.4, 2.7 \rangle.$$

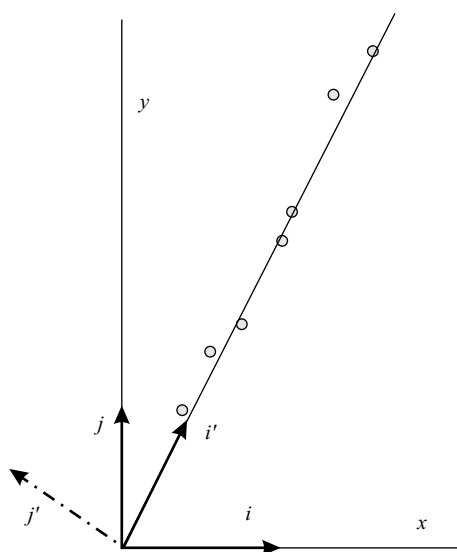
A főkomponensek meghatározása az alábbi MATLAB-parancsokkal történik:

- A bemeneti vektorok feltöltése:
`x1 = [1.0 1.8; 2.1 4.2; 0.7 1.6; 0.5 1.1; 1.3 2.5; 2.3 4.4; 1.8 3.5; 1.4 2.7]`
- Zérus várható érték biztosítása:
`x2 = x1 - ones(8,1)* mean(x1)`
- A PCA-módszer végrehajtása:
`[m1 m2 m3] = princomp(x2),`
ahol
 - `m1`: a sajátvektorok mátrixa,
 - `m2`: a vektorok koordinátái a sajátvektorok bázisában,
 - `m3`: a sajátértékek vektora.

Az eredményül kapott értékek:

```
m1 =  
 0.4679    0.8838  
 0.8838   -0.4679  
m2 =  
-0.9988    0.0903  
 1.6370   -0.0605  
-1.3159   -0.0812  
-1.8514   -0.0240  
-0.2398    0.0279  
 1.9073    0.0227  
 0.8779    0.0019  
-0.0162    0.0227  
m3 =  
 1.9004  
 0.0030
```

Az eredményből kiolvasható, hogy a PCA-elemzés egy olyan új bázisrendszert készített, amelyben a második dimenzió irányában a koordinátaértékek zérushoz közeli. Ezen dimenziót elhagyva a pontok egymás közötti távolságviszonya nem változik számottevően. A második dimenzió szerepének kicsiny jelentőségét a hozzá tartozó sajátérték igen kis értéke is jól mutatja.



1. ábra. Főkomponens-irányokra történő áttérés

Az 1. ábrán látható, hogy az új (i', j') koordináta-rendszerben az i' bázisvektor irányában jelentős, míg a j' bázisvektor irányában igen kicsi a koordinátaértékek varianciája. A j' dimenziót elhagyva tehát kicsiny információvesztéssel csökkenthető az eredeti adathalmaz dimenziója.