Componentes principales

Walter Sosa Escudero
wsosa@udesa.edu.ar

26 de octubre de 2020

Dos problemas

Bienestar

- 19 dimensiones del bienestar en Argentina
- Ingreso, educacion, tipo de trabajo, paga en cuotas, etc.
- Realmente el bienestar tiene 19 dimensiones?
- En el otro extremo, existe una unica dimension y en realidad cada variable es una medicion distinta de esta dimension?

Examen

- Cuatro preguntas: 1) Conceptual, 2)Teorica, 3) Practica (conceptual), 4) Empirica
- Estamos preguntando cuatro veces lo mismo?
- Las preguntas extraen informacion de distintas ‘dimensiones’ de lo evaluado?
- Es justo resumir todo en el promedio? Como ponderar?

Dimensionalidad via componentes principales

$X \equiv (x_1, x_2, \ldots, x_n)$, matriz $N \times K$

Factor: $F = X\delta, \qquad \delta \in \Re^K$

Idea: resume las K variables en una sola (F).

Terminologia: los coeficientes de δ son los *loadings*: cuanto 'importa' cada x_s en el factor.

Dimensionalidad: resumir las K variables originales en unos pocos $q < K$ factores.

Un poquito de algebra

- $A_{m\times m}$. Escalar λ tal que $Ax = \lambda x$ para un vector $m \times 1$, $x \neq 0$ es un *autovalor* de A. x es un *autovector* de A correspondiente al autovalor λ.
- $A_{m\times m}$ con autovalores $\lambda_1, \lambda_2, \ldots, \lambda_m$, entonces:
  (a) $\mathrm{tr}(A) = \sum_{i=1}^{m} \lambda_i$, (b) $|A| = \prod_{i=1}^{m} \lambda_i$
- Si $A_{m\times m}$ tiene m autovalores diferentes, entonces los autovectores asociados son todos linealmente independientes.
- Descomposicion espectral: $A = P\Lambda P'$, en donde $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_2)$ y P es la matriz cuyas columnas son los autovectores correspondientes.

Factores via componentes principales

- $x_1, x_2, \ldots, x_K$, K vectores de N observaciones cada uno.
- Factor: $F = X\delta$
- Cual es la 'mejor' combinacion lineal de $x_1, \ldots, x_K$?
- Mejor? Maxima varianza. Intuicion? La que mejor reproduce la variabilidad original de todas las $x's$

- $X = [x_1, \ldots, x_K]_{N \times K}$, $\Sigma = V(X)$, $\delta \in \Re^K$.
- $F = X\delta$ es una combinacion lineal de X, con $V(X\delta) = \delta'\Sigma\delta$.
- $\max_\delta \ \delta'\Sigma\delta$. Es obvio que la solucion pasa por llevar δ a infinito. Debemos *normalizar* δ.
- $\max_\delta \ \delta'\Sigma\delta$, sujeto a $\delta'\delta = 1$. Llamemos δ^* a la solucion a este problema.

$F^* = X\delta^*$ es la 'mejor' combinacion lineal de X. Es el primer componente principal.

Resultado: δ^* es el autovector correspondiente al mayor autovalor de $\Sigma = V(X)$.

$F^* = X\delta^*$ es el *primer componente principal* de X.

Intuicion: X tiene K columnas y $Y = X\delta$ tiene una sola. El factor construido con el primer componente principal es la mejor forma de representar las K variables de X usando una sola variable sola.

Solucion al problema del primer componente principal

Problema: $\max_\delta \ \delta'\Sigma\delta, \quad$ sujeto a $\delta'\delta = 1$

Lagrange: $L(\delta, \lambda) = \delta'\Sigma\delta + \lambda\,(1 - \delta'\delta)$. CPO:

$$\Sigma\,\delta = \lambda\,\delta$$

En el optimo, δ es el autovector correspondiente al autovalor λ de Σ. Premultiplicando por δ' y recordando que $\delta'\delta = 1$:

$$\delta'\Sigma\delta = \lambda$$

A fines de maximizar $\delta'\Sigma\delta$ debemos elegir λ igual al maximo autovalor de Σ y δ igual al autovalor correspondiente.

Factores como aprendizaje no supervisado

Regresion $Y = X\beta + u$. $\hat{Y} \equiv X\hat{\beta}$. Min $\sum(Y_i - \hat{Y})^2$

El aprendizaje es supervisado: la discrepancia entre Y y $\hat{Y}$ 'guia' el aprendizaje.

El problema de construccion de factores es no supervisado: construimos un indice (el factor) sin nunca verlo.

Empezamos con $X_{N\times K}$ y terminamos con $F^*_{N\times 1} = X\delta^*$

q componentes principales

- El primer componente principal? Hay otros?
- Consideremos el siguiente problema:

$$\max_{\delta_2} \;\; \delta_2' \Sigma \delta_2$$

$$\text{sujeto a: } 1) \;\; \delta_2' \delta_2 = 1, \quad 2) \; Cov(\delta_2' X, \delta^{*'} X) = 0$$

- $F_2^* = X\delta_2^*$ es el segundo componente principal: la mejor combinacion lineal que es ortogonal a la mejor inicial.
- Recursivamente, utilizando esta logica se pueden formar q componentes principales. Notar que algebraicamente podriamos construir $q = K$ factores.

- Sean $\lambda_1, \ldots, \lambda_K$ los autovalores de $\Sigma = V(X)$, ordenados de mayor a menor, y $p_1, \ldots, p_K$ los autovectores correspondientes. Llamemos P a la matriz de autovectores.
- *Resultado:* $\delta_j = p_j, \;\; j = 1, \ldots K$ ('loadings' de los componentes principales = autovectores ordenados de Σ).
- Sea $F_j = X\delta_j, \;\; j = 1, \ldots K$ el $j-$esimo componente principal. Es facil ver que

$$V(F_j) = \delta_j' \Sigma \delta_j = p_j' P \Lambda P' p_j = \lambda_j$$

(la varianza del $j-$ esimo componente principal es el $j-$ esimo autovalor ordenado de Σ).

Importancia relativa de los factores

La varianza total de X es la suma de las varianzas de $x_j, j = 1, \ldots, K$, o sea $\text{tr}\ \Sigma$. Es facil mostrar que

$$\text{tr}\ \Sigma = \text{tr}\ P\Lambda P' = \text{tr}\ \Lambda P P' = \sum_{j=1}^{K} \lambda_j = \sum_{j=1}^{K} V(F_j)$$

Entonces:

$$\frac{\lambda_k}{\sum_{j=1}^{K} \lambda_j}$$

mide la *importancia relativa* del j esimo componente principal.

Resultados Utiles

Sean $F_j = X\delta_j$, $j = 1, \ldots, N$ todos los componentes principales de X, cuya varianza es Σ. Entonces:

$$\rho_{F_j, x_k} = \delta_{kj} \sqrt{\frac{\lambda_j}{\sigma_{jj}}}$$

Intuicion: los coeficientes de las combinaciones lineales optimas (componentes principales) se relacionan con la correlacion entre cada variable y el componente principal.

Tips

- Mirar la importancia de los primeros componentes principales. Si el primero explica mucho, hay realmente una dimension sola (una dimension explica casi todo).
- Los coeficientes de los autovectores son ponderadores. Ver como cada una de las variables ‘participa’ en cada coeficiente principal.
- Cuidado con las diferencias de escala. Estandarizar siemmpre

Seleccion de factores

Supongamos que las columnas de X fueron estandarizadas, de modo de que cada variable tiene varianza unitaria. En este caso:

$$\text{tr } \Sigma = \sum_{i=1}^{K} V(x_i) = K$$

Y recordando que tr $\Sigma = \sum_{j=1}^{K} \lambda_j = \sum_{j=1}^{K} V(F_j)$, entonces:

$$\sum_{j=1}^{K} \lambda_j = K$$

Idea: en promedio, cada factor contribuye en una unidad. Cuando $\lambda_j > 1$, ese factor explica la varianza total mas que el promedio. Retener los factores con $\lambda_j > 1$

Interpretacion de factores?

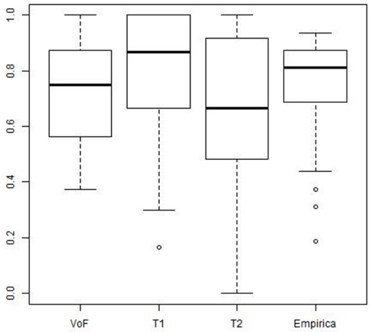
- $F_s = X\delta_s$: los 'loadings' muchas veces sugieren que un factor funciona como un 'indice' de un grupo de variables.
- Idea: mirar los 'loadings'
- Cuidado: los factores via componentes principales son ortogonales en forma recursiva.

Ejemplo: examen

- Calificaciones de un examen de cuatro preguntas. Las notas van de 0 a 1.
- Objetivos:
  - Garantizar ecuanimidad.
  - Justificar el uso del promedio como calificacion final.
  - Mostrar que “nada sobra”

```
. sum

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        var1 |        52    .7247596    .1934353       .375          1
        var2 |        52    .8102564    .2217516   .1666667          1
        var3 |        52    .6628205    .2722979          0          1
        var4 |        52    .7463942    .1756524      .1875      .9375
```



```
. cor
(obs=52)
```

	var1	var2	var3	var4
var1	1.0000			
var2	0.1995	1.0000		
var3	0.3083	0.4315	1.0000	
var4	0.3309	0.3901	0.5079	1.0000

. pca var1 var2 var3 var4

Principal components/correlation

Number of obs = 52
Number of comp. = 4
Trace = 4
Rho = 1.0000

Rotation: (unrotated = principal)

Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	2.10404	1.28885	0.5260	0.5260
Comp2	.81519	.221686	0.2038	0.7298
Comp3	.593504	.106237	0.1484	0.8782
Comp4	.487267	.	0.1218	1.0000

Principal components (eigenvectors)

Variable	Comp1	Comp2	Comp3	Comp4	Unexplained
var1	0.4051	0.8460	0.3459	-0.0223	0
var2	0.4821	-0.5111	0.6953	0.1514	0
var3	0.5527	-0.1496	-0.3297	-0.7506	0
var4	0.5460	-0.0250	-0.5368	0.6427	0