

UNIVERSIDAD NACIONAL DE COLOMBIA
DEPARTAMENTO DE ESTADÍSTICA
PCA

I. FUNDAMENTACIÓN

1. Explique en que consiste el método de los componentes principales.
2. Muestre que los componentes principales son combinaciones lineales de las variables que componen un vector aleatorio.
3. Muestre que la varianza total de un vector aleatorio es la suma de los valores propios de la matriz de covarianzas de dicho vector aleatorio.
4. Muestre que el valor esperado de los componentes principales es cero.
5. Muestre que la varianza del i -ésimo componente principal es el i -ésimo valor propio de la matriz de covarianzas del vector aleatorio original.
6. Explique cómo se puede medir la importancia del k -ésimo variable componente del vector aleatorio en el i -ésimo componente principal.
7. Muestre que la distancia de Mahalanobis al cuadrado se puede expresar como la distancia euclidiana de los componentes principales estandarizados al cuadrado.
8. Si cada variable componente de un vector aleatorio es estandarizada, muestre que el vector conformado por estas variables estandarizadas tiene media cero y matriz de covarianza igual a la matriz de correlación del vector aleatorio original.

II. EJERCICIOS

1. Del libro de Johnson & Wichern(2013) resolver los siguientes ejercicios
8.1-8.10, 8.12, 8.25

III. LECTURAS

Lea las siguientes dos referencias

Zou, H., Hastie, T., & Tibshirani, R. (2006). Sparse principal component analysis. *Journal of computational and graphical statistics*, 15(2), 265-286.

Hubert, M., Rousseeuw, P. J., and Vanden Branden, K. (2005), "ROBPCA: A New Approach to Robust Principal Component Analysis," *Technometrics*, 47, 64–79.

De acuerdo con estas lecturas:

1. Cuáles son las principales deficiencias del PCA clásico?
2. Qué deficiencia resuelve cada paper, y en qué consisten estas propuestas?
3. Para cada propuesta implemente un ejemplo que ilustre la metodología.
Compare los resultados obtenidos Vs la metodología clásica del PCA.