

UNIVERSIDAD NACIONAL DE COLOMBIA
DEPARTAMENTO DE ESTADÍSTICA
Análisis de Datos Funcionales
Taller 2: FPCA

FECHA DE ENTREGA: Febrero 3 de 2025

1. CUESTIONARIO

1. Con relación a los operadores lineales y al espacio de las funciones cuadrado integrables:
 - a) Explique por qué el espacio de las funciones continuas no es un espacio adecuado para representar las observaciones funcionales.
 - b) Explique por qué la igualdad en el espacio L^2 es en el sentido de casi siempre.
 - c) Qué propiedades tiene el operador de covarianza?
 - d) De acuerdo a las propiedades que presenta el operador de covarianza, presente una descomposición adecuada de este operador.
 - e) Sea Ψ un operador integral con el kernel gaussiano $\psi(t, s) = \alpha \exp(-(t^2 + s^2)/2)$.
 - 1) Para que valores de α Ψ es un operador de Hilbert-Schmidt.
 - 2) Muestre que $\|\Psi\|_{\mathcal{L}} \leq \|\Psi\|_{\mathcal{HS}}$.
 - 3) Encuentre la norma del kernel.
 - f) Explique qué es una base óptima para representar funciones aleatorias y presente una.
 - g) Explique cómo se encuentran las funciones propias y los valores propios del operador de covarianza.
 - h) Explique computacionalmente cómo se encuentran las estimaciones de las funciones propias y de los valores propios del operador de covarianza?
2. Estudie el artículo de Galeano, Joseph & Lillo (2015) e implemente los procedimientos de clasificación kNN y centroide, usando el conjunto de datos Tecator y la función de Mahalanobis. Explique los resultados obtenidos.
3. Considere las curvas espectrométricas usadas en el proceso de producción de azúcar.
 - a) Seleccione uno de los 7 procesos que hacen parte del estudio espectrométrico, y con las observaciones muestrales del proceso seleccionado:
 - 1) Verifique que los supuestos del FPCA se cumplen
 - 2) Encuentre las tres primeras funciones propias estimadas y sus valores propios asociados.
 - 3) Grafique la función de covarianza, junto con las tres primeras funciones propias. Explique esta gráfica.
 - 4) Cuántas funciones propias explican el 90 % de la varianza total del proceso?
 - 5) Expresé las tres primeras observaciones muestrales usando la expresión de Karhunen-Loève
 - 6) Grafique todas las realizaciones del primer score. Qué observa?
 - 7) Presente una matriz con los gráficos de dispersión de las realizaciones de los scores que conjuntamente explican el 90 % de la variación del proceso. Explique el comportamiento de los scores
 - 8) Expresé la función de covarianza en términos de los valores y funciones propias del operador de covarianza. Grafique algunas observaciones usando esta descomposición.

- b) Considere todo los 7 procesos espectrométricos del proceso de producción de azúcar:
- 1) Identifique y retire outliers multivariados.
 - 2) Usando el artículo de Berrendero, J. R., Justel, A., & Svarc, M. (2011), replique la sección 5.2 del artículo.
 - 3) Usando el artículo de Berrendero, J. R., Justel, A., & Svarc, M. (2011)
 - Encuentre las estimaciones de las tres primeras funciones propias multivariadas, gráfíquelas y presente una explicación de ellas.
 - Grafique Σ_{11} y Σ_{12} .
 - Cuántas funciones propias del operador de covaianza multivariado explican el 90 % de la varianza total del proceso. Gráfíquelas.
 - Grafique los valores propios del MFPCA Vs el porcentaje de varianza explicada.
 - Presente una tabla con las correlaciones de los scores asociados a los dos primeros procesos.
 - Exprese las tres primeras observaciones muestrales multivariadas usando la expresión de Karhunen-Loève
 - 4) Usando el artículo de Jacques, J., & Preda, C. (2014) resuelva los sub-items del item anterior. Compare los resultados obtenidos entre las metodologías propuestas por Jacques, J., & Preda, C. (2014) y Berrendero, J. R., Justel, A., & Svarc, M. (2011).
 - 5) Usando el artículo de Jacques, J., & Preda, C. (2014) encuentre los clusters asociados al dataset considerado
4. Usando los artículos de Happ, C., & Greven, S. (2018) y Happ-Kurz, C.(2020):
- a) Replique la sección 4.2: "Multivariate Functional Data Consisting of Functions and Images".
 - Genere el proceso multidimensional constituido por imágenes y curvas.
 - Usando el artículo de Rousseeuw, P. J., Raymaekers, J., & Hubert, M. (2018), encuentre imágenes outliers y retírelas del dataset.
 - Usando el artículo de Allen, G. I. (2013), encuentre las eigenimages y sus respectivos valores propios
 - Encuentre las funciones propias y valores propios de las curvas
 - Usando la metodología propuesta por Happ, C., & Greven, S. (2018, encuentre las funciones propias multidimensionales y sus respectivos valores propios.
 - b) Replique la sección 5 "Application-ADNI Study".

2. REFERENCIAS

- Allen, G. I. (2013, December). Multi-way functional principal components analysis. In 2013 5th IEEE international workshop on computational advances in multi-sensor adaptive processing (CAMSAP) (pp. 220-223). IEEE.
- Berrendero, J. R., Justel, A., & Svarc, M. (2011). Principal components for multivariate functional data. *Computational Statistics & Data Analysis*, 55(9), 2619-2634.
- Galeano, P., Joseph, E., & Lillo, R. E. (2015). The Mahalanobis distance for functional data with applications to classification. *Technometrics*, 57(2), 281-291.
- Happ, C., & Greven, S. (2018). Multivariate functional principal component analysis for data observed on different (dimensional) domains. *Journal of the American Statistical Association*, 113(522), 649-659.
- Happ-Kurz, C.(2020). Object-Oriented Software for Functional Data. *Journal of Statistical Software*, 93(5), 1–38. <https://doi.org/10.18637/jss.v093.i05>.
- Jacques, J., & Preda, C. (2014). Model-based clustering for multivariate functional data. *Computational Statistics & Data Analysis*, 71, 92-106.