
TRABAJO DE FIN DE GRADO EN MATEMÁTICAS

Departamento de Matemáticas
Universidad Autónoma de Madrid
(Curso académico 2020-21)

Título del proyecto: Análisis de datos funcionales: aproximación al problema de clasificación.

Nombre y apellidos: Pedro Martín Rodríguez-Ponga Eyriès

Nombre de los tutores: Luis Alberto Rodríguez Ramírez y José Luis Torrecilla Noguerales

INFORME INTERMEDIO

1.- Labor desarrollada hasta la fecha:

En el mes de septiembre comenzaron las reuniones con los tutores que han sido o bien hebdomadarias o bien quincenales. Las reuniones se han enfocado principalmente en lo siguiente:

- Introducción al análisis de datos funcionales.
- Recomendación de bibliografía.
- Resolución de dudas surgidas en la lectura.
- Verificar la asimilación de conceptos.
- Proporcionar una idea más intuitiva de algunos temas ilustrándolos mediante ejemplos más concretos.

Por otro lado, se ha mantenido una reunión adicional todas las semanas para tratar el proyecto paralelo de ampliación de una librería de Python [5] de análisis de datos funcionales.

2.- Esquema de los distintos apartados del trabajo:

Una estructura orientativa del trabajo podría ser:

- 0. Introducción al análisis de datos funcionales.
 - 0.1 Definición y motivación.
 - 0.2 Dificultades.
- 1. Introducción al problema de clasificación.
 - 1.1 Clasificador de Bayes: definición y optimalidad.
 - 1.2 Clasificación con datos funcionales.
- 2. Clasificación basada en profundidades.
 - 2.1 Definición y propiedades básicas.
 - 2.2 Profundidades en \mathbb{R} , \mathbb{R}^d , espacios funcionales.
 - 2.3 Maximum Depth, GG-Classifer, ...
 - 2.4 Comparativa y simulación.
- 3. Clasificación basada en distancias.
 - 3.1 Distancia a las medias recortadas.
 - 3.2 La distancia de Mahalanobis.
 - 3.3 Otras distancias.
 - 3.4 Comparativa y simulación.
- 4. Regresión logística.

3.- Descripción del proyecto:

La motivación del proyecto es el estudio de la extensión del problema de clasificación supervisada a un espacio funcional que puede tener infinitas dimensiones. Esta extensión permite trabajar con nuevos tipos de datos pero también añade nuevas dificultades como por ejemplo que las normas no son equivalentes. Concretamente, parte del trabajo se centrará en el estudio de como la elección de una métrica (o semi-métrica) afecta a los resultados de clasificación.

Entre los principales resultados se espera realizar una simulación ilustrando los diferentes métodos de clasificación, sus similitudes y sus diferencias.

Además, los diferentes métodos se implementarán en Python para poder aplicarlos a conjuntos de datos meteorológicos, médicos, poblacionales...

4.- Grado de consecución de los objetivos y plan de trabajo a desarrollar en la segunda mitad del periodo:

En la primera parte del curso se ha realizado un estudio de las principales referencias [1, 2] centrándose en la teoría general del análisis de datos funcionales así como en el problema de clasificación supervisada. Hasta la fecha el mayor esfuerzo ha sido en métodos de clasificación basados en profundidades.

Por otro lado, se ha estudiado análisis funcional de cara a aprovechar algunas propiedades de espacios funcionales dotados de un producto escalar como L^2 [12].

En la segunda mitad del periodo se pretende centrarse en clasificación basada en distancias insistiendo en la distancia de Mahalanobis. En caso de permitirlo el tiempo, se finalizaría con la regresión logística.

Referencias

- [1] RAMSAY, J. AND SILVERMAN, B. W. (2005.) Functional Data Analysis. Springer Series in Statistics. Springer.
- [2] FERRATY, F. AND VIEU, P. (2006.) Nonparametric Functional Data Analysis. Springer Series in Statistics. Springer.
- [3] TORRECILLA, J. L. (2015.) On the Theory and Practice of Variable Selection for Functional Data [Doctoral dissertation]. Universidad Autónoma de Madrid, España.
- [4] CUEVAS, A. (2013). A partial overview of the theory of statistics with functional data. Journal of Statistical Planning and Inference. <http://dx.doi.org/10.1016/j.jspi.2013.04.002>
- [5] GAA-UAM. (2021). scikit-fda. <https://github.com/GAA-UAM/scikit-fda>.
- [6] HUBERT, M., ROUSSEEUW, P. J., AND SEGAERT, P. (2017). Multivariate and functional classification using depth and distance. Advances in Data Analysis and Classification. 11. 445–466. 10.1007/s11634-016-0269-3.
- [7] FRAIMAN, R. AND MUNIZ, G. (2001). Trimmed means for functional data. Test, 10, 419–440.
- [8] PINTADO, S., AND ROMO, J. (2009). On the Concept of Depth for Functional Data. Journal of the American Statistical Association. 104. 10.1198/jasa.2009.0108.
- [9] GHOSH, A. K. AND CHAUDHURI, P. (2005b). On maximum depth and related classifiers. Scandinavian Journal of Statistics, 32, 327–350.

- [10] CUESTA-ALBERTOS, J. A., FEBRERO-BANDE, M. AND OVIEDO DE LA FUENTE, M. (2017). The DDG-classifier in the functional setting. *TEST*, 26. 119-142.
- [11] LI, J., CUESTA-ALBERTOS, J. A., AND LIU, R. Y. (2012). DD-classifier: Nonparametric classification procedure based on DD-plot. *Journal of the American Statistical Association*, 107(498):737-753.
- [12] RUDIN, W. (1974). *Real and complex analysis*, 2nd ed. McGraw-Hill, Inc., USA.