



LICENCIATURA EN CIENCIAS DE DATOS
Propuesta de tema de tesis

Detección de bloques de Independencia.

El objetivo de esta plan es explorar diferentes propuestas para detectar bloques de independencia de un vector aleatorio. Es decir, estimar cuales son los subvectores cuyas distribuciones factorizan la conjunta. Esta es una herramienta fundamental para reducir dimensi'on, siendo que bloques independientes se modelan de manera independiente.

Palabras clave: Independencia, selección de modelos, penalización

Conocimientos deseables

Segundo Ciclo completo de la carrera

¿Qué podría aprender quien realice esta tesis?

Modelos de Penalización para selección de modelos.

Dirección de la tesis

*Sued, Mariela
Instituto de Cálculo, FCEN, UBA. CONICET.*

Contacto: marielasued@gmail.com

Más información en el pdf a continuación.

Proyecto de tesis - Licenciatura en Ciencias de Datos

Título: *Detección de bloques/particiones de Independencia.*

El objetivo de esta plan es explorar diferentes propuestas para detectar bloques de independencia de un vector aleatorio. Es decir, estimar cuales son los subvectores cuyas distribuciones factorizan la conjunta. Esta es una herramienta fundamental para reducir dimensión, siendo que bloques independientes se modelan de manera independiente.

Palabras clave: Independencia, selección de modelo, penalización.

Sea $\mathbf{X} = (X_1, \dots, X_p)$ un vector aleatorio en \mathbb{R}^p , con distribución conjunta F : $\mathbf{X} \sim F$. Supongamos que existe $u_1 < u_2 < \dots < u_k$, coordenadas que permiten factorizar al vector en bloques independientes. En bla bla hicimos una primera propuesta para estimar el conjunto maximal que factoriza un vector aleatorio. Bajo normalidad, hay una propuesta alternativa basada en otras nociones de divergencia desarrollada por Yamila Alen durante su tesis de Licenciatura, bajo la dirección de Daniela Rodriguez. En esta tesis proponemos estudiar nuevas propuestas para abordar este mismo problema. Por un lado, explorar nuevas medidas de discrepancia, y por otro, aplicar algoritmos exitosos en otros contextos a este problema específico. A modo de ejemplo, pensamos en discrepancias basadas en funciones características y en utilizar *algoritmos genéticos* en este escenario. El trabajo consiste, principalmente, en explorar estas propuestas computacionalmente, mediante un estudio de simulación, y aplicar la metodología en datos reales. Cabe mencionar que no se requieren conocimientos de series de tiempo para la ejecución de este plan de trabajo.

Referencias

- Leonardi, F., Lopez - Rosenfeld, M., Rodriguez, D., Severino, M. T., & Sued, M. (2021). Independent block identification in multivariate time series. *Journal of Time Series Analysis*, 42(1), 19-33.
- Castro, B. M., Lemes, R. B., Cesar, J., Hünemeier, T., & Leonardi, F. (2018). A model selection approach for multiple sequence segmentation and dimensionality reduction. *Journal of Multivariate Analysis*, 167, 319-330.
- Detección de puntos de independencia para modelos Gaussianos Multivariados. (2023) Tesis de Licenciatura en Matemática de Yamila Alen, dirigida por Daniela Rodriguez (momentáneamente disponible a demanda).