

1. Información del equipo pedagógico y horarios de atención a estudiantes

Profesor magistral: Juan Camilo Prieto
Correo electrónico: jc.prietoa@uniandes.edu.co
Horario de atención a estudiantes: Definir por correo
Lugar de atención a estudiantes: A convenir

2. Descripción del curso

Este curso está diseñado para introducir a los estudiantes de maestría en economía a los fundamentos y aplicaciones del aprendizaje no supervisado, una rama del aprendizaje automático que se caracteriza por la ausencia de una variable de respuesta observada. A diferencia de los métodos de aprendizaje supervisado, aquí no se dispone de una “guía” explícita para evaluar los resultados, ya que los modelos trabajan exclusivamente con las variables predictoras disponibles.

El aprendizaje no supervisado permite explorar patrones, estructuras y relaciones en los datos que no están previamente etiquetados o definidos, haciendo de estos métodos una herramienta valiosa en escenarios donde no existe una variable objetivo clara.

En complemento al curso previo de machine learning, este curso aborda aplicaciones en las que no hay una respuesta definida. Los temas incluyen reducción de dimensionalidad, técnicas de clustering, sistemas de recomendación y análisis geográfico, con énfasis en casos relevantes para economistas.

La evaluación del curso se basará en talleres prácticos semanales, diseñados para asegurar que los estudiantes apliquen los conceptos aprendidos a problemas reales. Al finalizar el curso, los estudiantes estarán capacitados para:

- Identificar situaciones en las que el aprendizaje no supervisado sea una metodología adecuada.
- Seleccionar, implementar y evaluar los modelos apropiados.
- Interpretar y comunicar los resultados de manera efectiva, destacando su relevancia para la toma de decisiones en economía.

Este curso combina teoría y práctica, proporcionando a los estudiantes herramientas esenciales para abordar problemas complejos y no estructurados en su carrera profesional.

3. Resultados de aprendizaje

1. Reconocer las características generales del aprendizaje no supervisado y las situaciones en las que es pertinente aplicarlo.
2. Seleccionar, implementar e interpretar métodos de reducción de dimensión como el Análisis de Componentes Principales (PCA) y la Descomposición en Valores Singulares (SVD).
3. Implementar y evaluar algoritmos de clustering como K-medias, K-medoides, clustering jerárquico y DBSCAN.
4. Entender el funcionamiento del procesamiento del lenguaje natural (PLN) y modelos de tópicos como la Asignación Latente de Dirichlet (LDA).
5. Implementar y aplicar técnicas de procesamiento del lenguaje natural (PLN) y modelos de tópicos, como la Asignación Latente de Dirichlet (LDA), para el análisis de corpus de texto.
6. Crear e implementar técnicas avanzadas de filtrado colaborativo y modelos basados en texto (sistemas de recomendación).
7. Comprender los conceptos de puntos calientes y técnicas como estimación de densidad kernel para analizar patrones espaciales.
8. Comprender y desarrollar los conceptos de puntos calientes y técnicas como estimación de densidad kernel para analizar patrones espaciales.
9. Desarrollar habilidades prácticas para construir, implementar y evaluar modelos en diversos contextos aplicados utilizando herramientas computacionales.
10. Interpretar los resultados de los modelos implementados para tomar decisiones informadas en problemas complejos.
11. Seleccionar las metodologías más adecuadas según el tipo de datos y el problema planteado.

4. Cronograma

Tema	Subtemas
Reducción de Dimensión I: Análisis de Componentes Principales	<ul style="list-style-type: none"> • Reconocer las características generales del aprendizaje no supervisado y las situaciones en las que es pertinente su uso. • Reconocer el concepto de reducción de dimensión no supervisada y las circunstancias en las que es pertinente aplicarlo. • Reconocer las características y el funcionamiento del análisis de componentes principales. • Construir e implementar el análisis de componentes principales. • Interpretar los resultados del análisis de componentes principales para determinar el número adecuado de componentes a utilizar en la resolución de un problema en particular.
Reducción de Dimensión II: Descomposición en Valores Singulares	<ul style="list-style-type: none"> • Reconocer las características y el funcionamiento de la descomposición en valores singulares. • Construir e implementar la descomposición en valores singulares. • Interpretar los resultados de la descomposición en valores singulares.
Clustering I: K-medias y K-	<ul style="list-style-type: none"> • Reconocer en qué situaciones es necesario agrupar datos de forma no supervisada y su importancia.

medoides	<ul style="list-style-type: none"> • Reconocer las características y el funcionamiento de los algoritmos de K-medias y K-medoides • Construir e implementar los algoritmos de K-medias y K-medoides • Evaluar e interpretar el desempeño de los algoritmos de K-medias y K-medoides.
Clustering II: Clustering Jerárquico y DBSCAN	<ul style="list-style-type: none"> • Reconocer las características y el funcionamiento de algoritmos de clustering jerárquico y DBSCAN. • Construir e implementar los algoritmos de clustering jerárquico y DBSCAN. • Evaluar e interpretar el desempeño de los algoritmos de clustering jerárquico y DBSCAN. • Definir el conjunto de algoritmos más apropiado en la resolución de un problema en particular.
Sistemas de Recomendación I	<ul style="list-style-type: none"> • Reconocer las características generales de los Sistemas de Recomendación. • Reconocer las características y el funcionamiento de algoritmos de filtrado colaborativo basado en usuarios. • Reconocer las características y el funcionamiento del algoritmo de Análisis de Canasta de Compra. • Crear e implementar modelos de filtrado colaborativo basado en usuarios y de Análisis de Canasta de compra. • Evaluar e interpretar los resultados de filtrado colaborativo basado en usuarios y de Análisis de Canasta de compra.
Sistemas de Recomendación II	<ul style="list-style-type: none"> • Reconocer las características y el funcionamiento de las expresiones regulares • Reconocer las características y el funcionamiento de algoritmos de filtrado colaborativo basado en contenido. • Reconocer las características de los modelos de tópicos • Implementar expresiones regulares. • Crear e implementar modelos de filtrado colaborativo basado en contenido y modelo de asignación latente de Dirichlet. • Evaluar e interpretar los resultados de filtrado colaborativo basado en contenido y modelo de asignación latente de Dirichlet.
Análisis de Puntos Calientes Geográficos I	<ul style="list-style-type: none"> • Reconocer el concepto de puntos calientes, estimación de histogramas y estimación de densidad de kernel. • Identificar y organizar datos georreferenciados. • Crear e implementar modelos de estimación de densidad kernel bivariada para puntos calientes. • Interpretar los resultados de modelos de estimación de densidad kernel bivariada para determinar puntos calientes. • Construir e implementar el conjunto de algoritmos más apropiado en la resolución de un problema en particular.

<p>Análisis de Puntos Calientes Geográficos II</p>	<ul style="list-style-type: none"> • Reconocer las características y el funcionamiento del modelo de mezclas gaussianas • Crear e implementar modelos de mezclas gaussianas para el análisis de puntos calientes y clustering. • Evaluar e interpretar el desempeño y los resultados de modelos de mezclas gaussianas para el análisis de puntos calientes y clustering con datos georreferenciados.
--	---

5. Referencias

- Ahumada, H. A., Gabrielli, M. F., Herrera Gomez, M. H., & Sosa Escudero, W. (2018). Una nueva econometría: Automatización, big data, econometría espacial y estructural.
- Amat Rodrigo, Joaquín (2022) Ajuste de distribuciones con kernel density estimation y Python. Disponible en <https://www.cienciadedatos.net/documentos/pystats02-ajuste-distribuciones-kde-python.html>.
- Amat Rodrigo, Joaquín (2022). Clustering con Python. Available under Attribution 4.0 International (CC BY 4.0) at <https://www.cienciadedatos.net/documentos/py20-clustering-con-python.html>.
- Amat Rodrigo, Joaquín (2022). Clustering y heatmaps: aprendizaje no supervisado. Available under Attribution 4.0 International (CC BY 4.0) at https://www.cienciadedatos.net/documentos/37_clustering_y_heatmaps.
- Amat Rodrigo, J. Detección de anomalías con Gaussian Mixture Models (GMM) y python. Disponible en <https://www.cienciadedatos.net/documentos/py23-deteccion-anomalias-gmm-python.html>.
- Amat Rodrigo, Joaquín. (2018). Reglas de asociación y algoritmo Apriori con R, available under a Attribution 4.0 International (CC BY 4.0) at https://www.cienciadedatos.net/documentos/43_reglas_de_asociacion.
- Azzalini, A. and Bowman, A. W. (1990). A look at some data on the Old Faithful geyser. Applied Statistics, 39, 357–365. doi: 10.2307/2347385.
- Banik, R. (2018). Hands-on recommendation systems with Python: start building powerful and personalized, recommendation engines with Python. Packt Publishing Ltd.
- Bird, S., Klein, E., & Loper, E. (2009). Natural language processing with Python: analyzing text with the natural language toolkit. " O'Reilly Media, Inc."
- Bishop, Christopher M. Pattern Recognition and Machine Learning. (2006). Springer-Verlag Berlin, Heidelberg.

- Bivand, R. S., Pebesma, E. J., Gómez-Rubio, V., & Pebesma, E. J. (2008). Applied spatial data analysis with R (Vol. 747248717, pp. 237-268). New York: Springer.
- Cottam, J., Lumsdaine, A., & Wang, P. (2014). Abstract rendering: Out-of-core rendering for information visualization. Proceedings of SPIE The International Society for Optical Engineering. 9017. 90170K. 10.1117/12.2041200
- Google developers. (n.d.). Recommendation systems. Google. Disponible en <https://developers.google.com/machine-learning/recommendation/overview>
- Hartmann, K., Krois, J., Waske, B. (2018): E-Learning Project SOGA: Statistics and Geospatial Data Analysis. Department of Earth Sciences, Freie Universitaet Berlin.
- Patel, A. A. (2019). Hands-on unsupervised learning using Python: how to build applied machine learning solutions from unlabeled data. O'Reilly Media.

6. Metodología

El curso contiene distintas actividades diseñadas para cumplir con los objetivos de aprendizaje generales del curso y particulares propuestos en cada una de las semanas. De esta forma cada semana contiene la clase magistral donde abordaremos los problemas de la semana más teórica con implementaciones en python, scripts interactivos sobre los temas vistos en la clase y talleres que le permitirán abordar y aprender los temas vistos en la clase (cada dos semanas). Además, se espera que los alumnos realicen los ejercicios vistos en clase para afianzar los conocimientos.

Este es un curso de dos créditos, y por lo tanto exige una dedicación estimada de 12 horas semanales. Sin embargo, este puede variar y el tiempo reflejado en las actividades en la plataforma es menor y estimativo.

7. Evaluaciones

Con el fin de evaluar el aprendizaje de los estudiantes, se realizarán cuatro talleres en grupo, donde se aplicarán los temas vistos en las clases de manera práctica. Cada taller debe entregarse en grupos de 2 o 3 personas sin excepción. Las calificaciones se repartirán de la siguiente forma:

- Taller 1 - Reducción de Dimensión: 25% (6 de abril)
- Taller 2 - Clustering: 25% (20 de abril)
- Taller 3 - Sistemas de Recomendación: 25% (4 de mayo)
- Taller 4 - Modelos geográficos: 25% (18 de mayo)

Está permitido compartir código entre los compañeros de grupo para fines educativos y para facilitar la comprensión del problema a resolver, pero el análisis y las conclusiones deben ser propias. La universidad toma muy en serio el fraude académico y no será tolerado.

8. Asistencia

No hay control de asistencia durante las clases. Sin embargo, los estudiantes son responsables de enterarse de todo lo que se diga durante la clase, aunque esto no se encuentre en el programa del curso.

De acuerdo con el Artículo 45 del reglamento general de estudiantes de maestría, si hay inasistencias a una evaluación, los estudiantes tendrán tres (3) días calendario para presentar una excusa válida. Todas las excusas, sin excepción, serán validadas para corroborar que se ajustan al reglamento oficial de la universidad. Para una descripción de lo que constituye una incapacidad, ver reglamentación de las incapacidades estudiantiles.

9. Políticas generales de los cursos de Economía y fechas importantes

Los estudiantes deben consultar [este enlace](#), donde se encuentran las reglas sobre asistencia a clase, excusas válidas, fraude académico y faltas disciplinarias, reclamos, políticas de bienestar y fechas importantes del semestre.