

CURSO DE POSGRADO

Reconocimiento de Patrones y Aprendizaje de Máquina

Dr. Claudio Delrieux – DIEC – UNS

Carga horaria: 84 horas

Fundamentación

La reciente consolidación de las tecnologías y metodologías emergentes asociadas al aprendizaje de máquina están permitiendo el surgimiento de un amplio conjunto de aplicaciones e usos innovadores, tanto en el contexto científico-productivo como en las organizaciones gubernamentales y no gubernamentales. El presente curso propone familiarizar al alumno a los aspectos teóricos y tecnológicos fundamentales asociados al aprendizaje de máquina, así como temas relevantes asociados como la gestión y análisis de grandes datos. Se propone además el análisis de casos de estudio, y la realización de ejemplos de experiencias del ciclo completo del desarrollo de modelos de Ciencia de Datos.

Objetivos

En este curso se propone brindar al alumno la formación y experiencia en la gestión completa del ciclo de vida de una aplicación que requiera aprendizaje de máquina, así como el conocimiento de las plataformas de desarrollo existentes, la integración con desarrollo en la nube, redes sociales, etc. Al finalizar el curso los estudiantes serán capaces de implementar, integrar, o generar soluciones avanzadas en diversos campos de la ingeniería con la utilización de dichos sistemas y su integración con sistemas ya existentes. Reconocerán, a la vez, los fundamentos, las ventajas y prestaciones de las diferentes metodologías y modelos de aprendizaje de máquina, las plataformas y bibliotecas de software existentes, y la integración con diversas fuentes de información. Se capacitarán para el aprovechamiento de las herramientas de código abierto disponibles y la formación continua para poder asimilar los permanentes cambios tecnológicos asociados. Estos objetivos se cumplirán en forma incremental por medio de trabajos prácticos y un trabajo final integrador en los que se desarrollarán y resolverán problemas de aprendizaje de máquina con datasets reales, aplicando las técnicas abarcadas en el temario y su aplicación a la resolución de problemas concretos.

Contenidos

1. Introducción, motivaciones y conceptos básicos. Ingeniería de datos y modelos de gestión. Tipos y fuentes de datos. Fuentes de datos. Tecnologías de gestión de datos. Data wrangling y calidad de datos. Políticas de gestión.
2. Análisis de datos y analíticos. Minería de datos, aprendizaje de máquina y analíticos. Análisis de datos, exploración y confirmación. Modelos, productos de datos y analíticos. Atributos, valores y tipos de predictivos. Espacio de atributos. Modelos estadísticos. Estimadores multivariados, correlación.
3. Espacios de atributos. Distancia Euclídea, estandarizada, y de Mahalanobis. Normalización. Reducción de dimensionalidad. Componentes principales. Modelos no lineales. T-SNE. U-MAP.
4. Clasificación. K-NN. Clasificador Bayesiano. Regresión logística. Árboles de decisión. Random Forests. Sistemas basados en reglas. Máquinas de Vector Soporte. Perceptrones. Métricas de calidad de modelos de clasificación.
5. Clustering. Modelos basados en similitud. K-medias. Algoritmos aglomerativos y divisivos. Detección de atípicos. Evaluación de la calidad. Distorsión.
6. Regresión. Modelos univariados y multivariados. Modelos lineales y no lineales. Regularización. Evaluación de la calidad, error cuadrático medio, coeficiente de determinación.

7. Otros modelos de aprendizaje. Lenguaje natural. Aprendizaje de estructuras. Series de tiempo. Redes sociales.
8. Visualización interactiva de información. El sistema perceptual y cognitivo humano. Atributos visuales y representación de información. Modelos de representación (datos tabulares, espaciales, grafos). Diseño de vistas. Primitivas de diseño. Vistas accionables. Vistas múltiples coordinadas, filtrado, agregación. Vistas navegables, selección de ítems.

Metodología

El curso consiste en clases teóricas (inicialmente a través de encuentros por plataformas virtuales), donde se presentan los fundamentos de cada uno de los temas propuestos y se introducen los elementos para la realización de los trabajos prácticos. Dependiendo del número de alumnos, la realización de los mismos se realizarán en comisiones de no más de dos alumnos. Estas comisiones trabajarán individualmente, supervisados constantemente por la cátedra a través del uso de plataformas colaborativas como GitHub y Colab. Durante el cursado se realizarán experiencias prácticas del uso de aprendizaje de máquina con datasets reales obtenidos en plataformas como Kaggle o VAST, proponiendo problemas de complejidad creciente y requiriendo los diversos modos de preparación y análisis de datos vistos en la materia. En estas experiencias se aplicarán los conocimientos teóricos para resolver problemas puntuales usando y/o combinando modelos a fin de conocer la temática de la materia en forma práctica. La aprobación será mediante el desarrollo de un trabajo final desarrollado individualmente o en comisiones de dos personas. Toda propiedad intelectual utilizada (código, métodos, etc.) debe contar con las referencias adecuadas. Se tiene en cuenta para la calificación (a) presentación general, (b) prolijidad, (c) claridad, (d) redacción y corrección, (e) citas en la redacción, (f) calidad de código fuente (comentarios, indentación, uso de variables y estructuras de programación, etc.), (g) interpretación de resultados, (h) conclusiones y ensayos realizados para soportarlas.

Bibliografía

- [1] Casella G, Berger RL. Statistical inference. Second edition. Delhi. Cengage Learning 2017.
- [2] Hastie T, Tibshirani R, Friedman JH. The elements of statistical learning: data mining, inference, and prediction. 2nd ed. New York. Springer 2009.
- [3] Murphy KP. Machine learning: a probabilistic perspective. Cambridge, Mass. MIT Press 2012.
- [4] Barber D. Bayesian reasoning and machine learning. Cambridge. Cambridge University Press 2012.
- [5] Hastie T, Tibshirani R, Friedman JH. The elements of statistical learning: data mining, inference, and prediction : with 200 full-color illustrations. New York. Springer 2001.
- [6] Goodfellow I, Bengio Y, Courville A. Deep learning. Cambridge, Massachusetts. The MIT Press 2016.
- [7] Raschka S. Python machine learning: unlock deeper insights into machine learning with this vital guide to cutting-edge predictive analytics. Birmingham. Packt Publishing 2015.
- [8] McKinney W. Python for data analysis. Farnham. O'Reilly 2013.
- [9] Harrington P. Machine learning in action. Shelter Island, N.Y.. Manning Publications 2012.
- [10] Wes McKinney. Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython (2nd ed.), O'Reilly 2018.
- [11] Big Data Fundamentals: Concepts, Drivers & Techniques. Thomas Erl, Wajid Khattak, Paul Buhler. Prentice Hall, 2012.
- [12] Mining of Massive Datasets. Jure Leskovek, Anand Rajaraman and Jeffrey Ullman. v2.1, Cambridge University Press. 2014.
- [13] Conway, Drew; White, John Myles (February 2012). Machine Learning for Hackers. O'Reilly Media. ISBN 978-1449303716.

- [14] Russel, Matthew A. (October 2013). Mining the Social Web, 2nd Edition. O'Reilly Media. ISBN 978-1449367619.
- [15] Ben Bederson and Ben Shneiderman (2003). The Craft of Information Visualization: Readings and Reflections. Morgan Kaufmann.
- [16] Riccardo Mazza (2009). Introduction to Information Visualization, Springer.
- [17] Colin Ware (2000). Information Visualization: Perception for design. San Francisco, CA: Morgan Kaufmann.
- [18] Doing Data Science. Rachel Schutt, Cathy O'Neil. O'Reilly Editors, 2014. ISBN: 978-1-449-35865-5