



Ministerio de Cultura y Educación
Universidad Nacional de San Luis
Facultad de Ciencias Físico Matemáticas y Naturales
Departamento: Informática
Area: Area V: Automatas y Lenguajes

(Programa del año 2019)

I - Oferta Académica

Materia	Carrera	Plan	Año	Período
(OPTATIVA) APRENDIZAJE AUTOMATICO Y MINERIA DE DATOS	LIC.CS.COMP.	32/12	2019	1° cuatrimestre

II - Equipo Docente

Docente	Función	Cargo	Dedicación
ERRECALDE, MARCELO LUIS	Prof. Responsable	P.Asoc Exc	40 Hs
FERRETTI, EDGARDO	Responsable de Práctico	JTP Exc	40 Hs

III - Características del Curso

Credito Horario Semanal				
Teórico/Práctico	Teóricas	Prácticas de Aula	Práct. de lab/ camp/ Resid/ PIP, etc.	Total
Hs	2 Hs	2 Hs	1 Hs	5 Hs

Tipificación	Periodo
B - Teoria con prácticas de aula y laboratorio	1° Cuatrimestre

Duración			
Desde	Hasta	Cantidad de Semanas	Cantidad de Horas
13/03/2019	22/06/2019	15	75

IV - Fundamentación

A partir de la popularización de Internet y el incremento del número de bases de datos accesibles, se ha producido en los últimos años un gran crecimiento en los volúmenes y disponibilidad de datos alrededor del mundo. Se ha estimado que la cantidad de datos almacenados en las bases de datos del mundo se duplica cada veinte meses. Este crecimiento cuantitativo de los datos no se ha reflejado en un crecimiento cualitativo de la información disponible. Existe un "gap" creciente entre la generación de los datos y nuestro entendimiento de los mismos.

Una alternativa para solucionar este problema que ha recibido un interés creciente es el área de minería de datos. La minería de datos (en inglés Data Mining y de ahora en más MD) es el proceso de descubrir patrones en grandes volúmenes de datos. Este proceso debe ser automático (o más usualmente) semi-automático. Los patrones descubiertos deben ser significativos en el sentido que deben conducir a alguna ventaja (generalmente económica).

V - Objetivos / Resultados de Aprendizaje

Introducir al alumno en los principales conceptos vinculados al aprendizaje automático y la minería de datos. Al finalizar el curso los alumnos deberán ser capaces de identificar las principales etapas y procesos de la minería de datos, los mecanismos estadísticos y de aprendizaje automático frecuentemente utilizados en este área. Se espera además que los mismos sean

VI - Contenidos

Unidad 1. Aprendizaje Automático (AA).

Definición. Importancia y Aplicaciones del AA. Tarea de Aprendizaje bien definida. Tipos de feedback para el aprendiz. Diseño de un sistema de aprendizaje. Aspectos de estudio relevantes en AA.

Unidad 2. Aprendizaje Supervisado

Tipos de aprendizaje. Aprendizaje automático (supervisado). Aprendizaje de una hipótesis. Esquema General. Aprendizaje como Búsqueda. La necesidad del sesgo inductivo. Evaluación de una hipótesis. Medidas de evaluación de clasificadores.

Unidad 3. Minería de Datos. Aspectos Generales

Definición de Minería de Datos (MD). Origen. Relación con otras disciplinas. Áreas de aplicación. Aprendizaje de Máquina (AM). Etapas del proceso KDD. AM versus MD. MD en el contexto del proceso KDD. Principales tareas de la minería de datos. Clasificación, Aprendizaje de conceptos, Categorización, Regresión, Análisis de grupos (cluster analysis), Análisis de asociaciones (association analysis), Análisis de correlaciones.

Unidad 4: Aprendizaje de árboles de decisión.

Algoritmo ID3. Medidas de entropía y ganancia de información. Sesgo inductivo en ID3. Métodos para evitar el sobre-entrenamiento.

Unidad 5: Máquinas de vectores soporte (SVM).

Introducción. SVM para clasificación binaria. Justificación teórica. Aplicaciones de SVM. Extensiones y temas avanzados. Software para SVM.

Unidad 6: Regresión Lineal, descenso del gradiente y redes neuronales

Regresión Lineal (RL) Simple. RL múltiple. Mínimos cuadrados ordinarios. Descenso del gradiente. Redes Neuronales (RN) como una herramienta para hacer regresión no lineal. Estructura de las RN. Perceptrones. RN feed-forward con múltiples capas. Algoritmo de Back-Propagation.

Unidad 7. Redes Neuronales Profundas

Redes neuronales convolucionales. Primeros desarrollos: Neocognitron, Lenet 5. Operación y capas de convolución. Operación de pooling. Capa de dropout. Desarrollos recientes: VGG, Inception. Redes neuronales recurrentes. Ideas Iniciales, Tipos de secuencia. Unidades de memoria, Long Short Term Memory. Gated Recurring Units. Ejemplos de aplicación.

Unidad 8: Otras componentes del proceso de KDD.

Preparación de los datos. Evaluación, interpretación y visualización de patrones.

Unidad 9: Aplicaciones

Aspectos prácticos involucrados en el desarrollo de aplicaciones de MD para un problema particular. Caso de estudio: Minería de Textos y la Web. Herramientas de desarrollo de aplicaciones para MD. El sistema Weka. RapidMiner. Scikit-learn. Bases de datos públicas.

VII - Plan de Trabajos Prácticos

Práctico 1: Aspectos generales. Aprendizaje de conceptos.

Práctico 2: Aprendizaje de Árboles de decisión.

Práctico 3: SVM

Práctico 4: Redes Neuronales

Práctico 5: Desarrollo de un proyecto de programación a ser entregado a final de cuatrimestre.

Práctico de Máquina Nro. 1. Introducción al aprendizaje automático con Scikit-Learn.

Práctico de Máquina Nro.2. Aprendizaje de árboles de decisión

Práctico de Máquina Nro.3. SVM y regresión lineal

Práctico de Máquina Nro. 4. Redes Neuronales Artificiales

VIII - Regimen de Aprobación

La materia sólo admite el régimen PROMOCIONAL. En este contexto, la aprobación del curso requiere:

- a. La asistencia del alumno a al menos un 80% de las clases teóricas y prácticas.
- b. La entrega de la totalidad de los trabajos prácticos solicitados. Cada uno de estos trabajos, tendrá dos instancias de recuperación en caso de no alcanzarse una nota mínima de 7 puntos.
- c. Entrega de un proyecto de programación propuesto vinculado a las temáticas del curso, con un informe adicional con las principales conclusiones.
- d. Aprobar, con un mínimo de 7 (siete) puntos, un coloquio integrador oral y/o escrito al final del cuatrimestre.

La nota final se computará promediando las notas obtenidas en cada uno de los puntos mencionados previamente.

IX - Bibliografía Básica

- [1] "Introduction to Data Mining". Tan, Steinbach y Kumar. Pearson. 2006.
- [2] "Data Mining: Practical Machine Learning Tools and Techniques". Third Edition. Witten y Frank. Morgan Kaufmann. 2011.
- [3] "Introducción a la Minería de datos". Orallo, J. H., Ramírez Quintana, Ma. J. y Ramírez, C.F.. Pearson Prentice Hall. 2004.
- [4] "Machine Learning". Tom Mitchell. McGraw-Hill Series in Computer Science, 1997.
- [5] "Readings in Machine Learning". Shavlik, J. y Dietterich, Thomas. Morgan Kaufmann, 1990.
- [6] "The Elements of Statistical Learning". T. Hastie, R. Tibshirani, J. H. Friedman. Springer, 2002.
- [7] "Principles of Data Mining (Adaptive Computation and Machine Learning)". David J. Hand, Heikki Mannila, Padhraic Smyth. MIT Press, 2001.

X - Bibliografía Complementaria

- [1] "The Text Mining HandBook: Advanced Approaches in Analyzing Unstructured Data". Feldman and Sanger. Cambridge University Press. 2007.
- [2] "Deep Learning". Ian Goodfellow and Yoshua Bengio and Aaron Courville. Mit Press, 2016.
- [3] "Python Data Science Handbook: Essential tools for working with data". Jake VanderPlas (1ra. Edición). 2017.
- [4] "Introduction to Machine Learning with Python". Andreas C. Müller and Sarah Guido. (1ra. Edición). 2016
- [5] "Python Machine Learning". Sebastian Raschka. 2015.
- [6] "Python for Data Analysis". Wes McKinney. (2da. Edición). 2016
- [7] Publicaciones y apuntes complementarios sobre los distintos tópicos del curso que estarán disponibles en <http://www.dirinfo.unsl.edu.ar/~aamd>

XI - Resumen de Objetivos

Introducir al alumno en los principales conceptos vinculados al aprendizaje automático y la minería de datos. Al finalizar el curso los alumnos deberán ser capaces de identificar las principales etapas y procesos de la minería de datos y los mecanismos de aprendizaje automático frecuentemente utilizados en este área. Se espera además que los mismos sean capaces de detectar y aplicar las técnicas de minería de datos en aplicaciones concretas del mundo real.

XII - Resumen del Programa

Unidad 1: Introducción al Aprendizaje Automático.

Unidad 2: Aprendizaje Supervisado
Unidad 3: Introducción a la Minería de Datos.
Unidad 4: Aprendizaje de Árboles de Decisión.
Unidad 5: Máquinas de vectores soporte (SVM).
Unidad 6: Regresión Lineal, descenso del gradiente y redes neuronales
Unidad 7: Redes Neuronales Profundas.
Unidad 8: Otras componentes del proceso de KDD.
Unidad 9: Desarrollo de aplicaciones de Minería de Datos.

XIII - Imprevistos

XIV - Otros