

**UNIVERSIDAD DE PUERTO RICO
RECINTO UNIVERSITARIO DE MAYAGUEZ
DEPARTAMENTO DE MATEMATICAS**

COMP6839: MINERIA DE DATOS (DATA MINING)

Rev. Enero 12 del 2009

Instructor: Edgar Acuna

Objetivos del curso: Entender los conceptos fundamentales para llevar a cabo minería de datos y descubrimiento de conocimiento en base de datos. Experimentar algunos algoritmos más usados en minería de datos en conjuntos de datos reales.

Horario del curso: M y J de 9 a 10.15 am en M314.

Pre-requisitos del curso: Haber tomado dos cursos donde se hayan visto conceptos estadísticos. Tener algún conocimiento de matrices, sistemas de bases de datos y de algún programa de computación.

Lenguaje de programación: Programas gratuitos de de minería de datos tales como R, Weka, RapidMiner(Yale) y Orange. Microsoft SQL también puede ser usado.

Oficina: M314.

Horas de oficina: M y J de 7.30 a 9.00am , W de 8 a 11am.

Extension: x3287

Correo electronico del Profesor: eacuna@uprm.edu ,
edgar@math.uprm.edu, edgar@cs.uprm.edu eacunaf@yahoo.com

Pagina web del curso: math.uprm.edu/~edgar/comp6838.html

Ayudante: Roxana Aparicio (M309).

Contenido del curso

- I Clasificación Supervisada (18 horas)
 - Clasificación Bayesiana (2)
 - Precisión de un clasificador(1)
 - Clasificación usando “naïve Bayes” (2)
 - Clasificación usando análisis discriminante Lineal (2.5)
 - Clasificación usando arboles de decision (2.5)

Clasificación usando vecinos cercanos (k-nn) (2)
Clasificación usando regresión logística (2)
Clasificación usando redes neurales(2)
Combinacion de clasificadores: Bagging y Bosting (2)

EXAMEN I

- II. Analisis por conglomerados (15 horas)
 - Métodos de Particionamiento: El Algoritmo de k-medias y el PAM (4)
 - Métodos Jerárquicos: Algoterativos y divisivos (4)
 - Mapas auto-organizantes (SOM) de Kohonen (2)
 - Metodos de conglomerados basados en modelos(2)
 - Otros métodos de conglomerados (1)
 - Validación de conglomerados (2)

- III. Reglas de asociación (10 horas)
 - Analisis de la canasta de Mercado (1)
 - Generación de itemsets frecuentes (2)
 - El algoritmo apriori (2)
 - Evaluacion de patrones de asociación (2)
 - Manipulación de atributos categóricos y continuos (1)
 - Patrones secuenciales (2)

Texto:

Tan, P.N., Steinbach, M. y Kumar, V. (2006) Introduction to data Data Mining. Pearson Addison Wesley, Boston.

Han, J. y Kamber, M. (2006). Data Mining: Concepts and Techniques. Second Edition. Morgan Kaufman Publishers. San Francisco. La primera edición de este libro esta disponible gratis en la internet en <http://www.cs.wmich.edu/~yang/teach/cs595/han>

References

1. Hastie, T, Tibshirani, R & Friedman, J. (2001). *Elements of Statistical Learning*. Springer-Verlag.
2. Kantardzic M. (2003). *Data Mining: Concepts, Models, methods, and Algorithms*. John Wiley. New York.
3. Williams, G. [Data Mining Desktop Survival Guide](#), on-line book (PDF).
4. Witten, I.H., and Frank, E. (2005), *Data Mining: Practical Machine Learning tools and techniques*. Morgan Kaufman, San Francisco.

Evaluación

1	Tareas (4)	40%
2.	Examen Parcial	25%
3.	Proyecto incluyendo presentación oral	35%

LINKS

<http://www.cs.sfu.ca/~han/dmbook> Slides del libro de Han

<http://www.eas.asu.edu/~mining03>. Curso de Data Mining por el professor Liu.

<http://www.liacc.up.pt/~ltorgo/DataMiningWithR/>. Libro de Data Mining con R por Luis Torgo.

<http://www.kdnuggets.com/dmcourse> Curso en Data mining por kdnuggets

<http://www.galaxy.gmu.edu/stats/syllabi/DMLIST.html>. Varios links a Data Mining

http://www.pcc.qub.ac.uk/tec/courses/datamining/stu_notes/dm_book_1.html . Notas de Data Mining de la Queen's University de Belfast.

<http://www.scd.ucar.edu/hps/GROUPS/dm/dm.html> Varios links a Data Mining

<http://ocw.mit.edu/OcwWeb/Sloan-School-of-Management/15-062Data-MiningSpring2003/LectureNotes/>. Curso de Data Mining en MIT

<http://www.cise.ufl.edu/class/cis6930fa03dm/notes.html> Curso de Data Mining, CISE, University of Florida

<http://www.cs.uu.nl/docs/vakken/adm/>. Notas en data Mining del Profesor Felders

<http://www.cs.unc.edu/Courses/comp290-90-f04/> Notas en Data Mining del Profesor Wang

<http://www.cs.pdx.edu/~timm/dm/>. Notas en Data Mining de Pórtland University

<http://www.stats.ox.ac.uk/~vos/DataMining.html> Data Mining for Bioinformatics with R

<http://ljsavage.wharton.upenn.edu/~stine/mich/>. Notas de Data Mining de University of Pennsylvania.

<http://www.galaxy.gmu.edu/stats/syllabi/INFT979.spring2001.html> Notas en data mining del professor Wegman.