



Programa de MODELOS ESTADÍSTICOS PARA CLASIFICACIÓN Y REGRESIÓN

1. NOMBRE DE LA UNIDAD CURRICULAR

MODELOS ESTADÍSTICOS PARA CLASIFICACIÓN Y REGRESIÓN

2. CRÉDITOS

6 CRÉDITOS

3. OBJETIVOS DE LA UNIDAD CURRICULAR

Este curso es un primer curso de modelización estadística cuyo objetivo es de acercar el estudiante al manejo de base de datos con el fin de ser usadas para hacer inferencia o predicción. El mismo se basa fuertemente en el curso de Geometría y Álgebra Lineal para el manejo de las matrices de datos.

El curso proporciona los conocimientos necesarios para comprender e implementar algoritmos de análisis multivariado. Al terminarlo, el estudiante habrá adquirido la metodología básica para resolver algunos problemas prácticos de análisis de datos, clasificación y regresión y tener una base para continuar la formación en el área.

Se procurará que el estudiante sepa aplicar dichas técnicas a conjuntos de datos reales e interpretar los resultados obtenidos acercándolo al empleo de los paquetes para el análisis estadístico de datos disponibles en un ambiente de desarrollo de software libre.

4. METODOLOGÍA DE ENSEÑANZA

Curso teórico-práctico de 6 horas semanales, durante 8 semanas.

5. TEMARIO

Incluye una descripción general de los grandes temas del curso y de los subtemas incluidos en cada uno de ellos.

1. BASE DE DATOS. Variables explicativas y predictivas. Variables ordinales, categóricas y continuas. Repaso de álgebra matricial.
2. ANÁLISIS DE COMPONENTES PRINCIPALES (ACP). Cálculo de los componentes principales. Relaciones entre las nuevas y las viejas variables. Análisis Dual. Representación en el plano factorial de los individuos y de las variables.
3. CLASIFICACIÓN Y REGRESIÓN. Aprendizaje automático supervisado y no supervisado. Modelización: inferencia y predicción. Compromiso Sesgo-Varianza. Error de predicción.
4. REGRESIÓN:
 - a) Regresión lineal simple. Relación con los encares vistos en los cursos previos: normal bivariada, método de mínimos cuadrados en álgebra lineal y optimización.
 - b) Regresión lineal múltiple. Tests de significancia de los parámetros. ANOVA
 - c) Series temporales y modelo autorregresivo AR(1)
5. CLASIFICACIÓN:
 - a) Naive Bayes
 - b) k vecinos más cercanos,
 - c) Árboles de clasificación. Construcción. criterios de partición, de parada y de asignación. Algoritmo de poda. Elección del mejor subárbol. Regla 1-SE.

6. BIBLIOGRAFÍA

6.1 Básica

1. Análisis de datos multivariantes, Daniel Peña, Mac Graw Hill, 2002
2. An introduction to Statistical Learning with Applications in R, G.James, D. Witten, T. Hastie, R. Tibshirani, Springer, 2013.
3. Pattern Classification, Richard O Duda, Peter E Hart and David G Stork, John Wiley & Sons, 2012.
4. Análisis de series temporales, Daniel Peña, Alianza Editorial, 2010.

6.2 Complementaria

1. Modern Multivariate Statistical Techniques, Alan Julian Izenman, Springer, ISBN: 978-0-387-78188-4, 2008.
2. Multivariate Statistics: High-Dimensional and Large-Sample Approximations, Yasunori Fujikoshi, Vladimir V. Ulyanov, Ryoichi Shimizu, Wiley, ISBN: 978-0-470-41169-8, 2010.
3. Multivariate Statistical Inference and Applications, Alvin C. Rencher, Wiley, ISBN: 0-471-57151-2, 1998.
4. A First Course in Multivariate Statistics, Bernhard Flury, Springer, ISBN: 978-0-387-98206-9, 1997.
5. A Handbook of Statistical Analyses Using R, 2nd edition, Brian S. Everitt y Torsten Hothorn, Chapman & Hall/CRC, ISBN: 978-1-4200-7933-3, 2010.

7. CONOCIMIENTOS PREVIOS EXIGIDOS Y RECOMENDADOS

7.1 Conocimientos Previos Exigidos: son requeridos conocimientos básicos de probabilidad y estadística, como variables aleatorias, distribuciones, estimación puntual y por intervalos, ley de los grandes números y teorema central del límite. Se requiere también manejo fluido de matices, en especial los teoremas espectrales, así como de la geometría euclidiana (norma, distancia, producto escalar, proyecciones ortogonales). Finalmente se requiere conocimiento de maximización/minimización de funciones de varias variables.

7.2 Conocimientos Previos Recomendados: Programación en R, Excel o Octave.

No incluye la información de previaturas. Las unidades curriculares previas serán definidas por cada carrera que tome la unidad curricular y serán incluidas en el anexo B.

ANEXO A

Para todas las Carreras

Esta primera parte del anexo incluye aspectos complementarios que son generales de la unidad curricular.

A1) INSTITUTO

Indicar el Instituto a cargo de la unidad curricular.

A2) CRONOGRAMA TENTATIVO

Consiste en un cronograma de avance semanal con detalle de las horas de clase asignadas a cada tema.

Semana 1	Base de datos. Distinto tipo de variables
Semana 2	Análisis de Componentes Principales
Semana 3	Problema de clasificación y de regresión. Regresión lineal simple
Semana 4	Regresión lineal simple
Semana 5	Regresión lineal múltiple
Semana 6	Métodos de clasificación: Naive Bayes, k vecinos más cercanos
Semana 7	Arboles de clasificación
Semana 8	Series temporales y modelo AR(1)

A3) MODALIDAD DEL CURSO Y PROCEDIMIENTO DE EVALUACIÓN

Curso teórico-práctico.

La evaluación constará de una prueba escrita final presencial y de la entrega de un trabajo donde se aplicará las técnicas vistas en el curso a una base de datos elegida.

A4) CALIDAD DE LIBRE

Indicar si en la unidad curricular los estudiantes podrían acceder o no a la Calidad de Libre.

No posee calidad de libre.

A5) CUPOS DE LA UNIDAD CURRICULAR

(En caso de que corresponda, indicar los cupos totales.)

Cupos mínimos:

Cupos máximos:

Nota:

Si se definen cupos, en una nota aparte se deberá incluir:

- *motivo por el cual la unidad curricular tiene cupos (tanto máximos como mínimos).*
- *el mecanismo de selección para cuando se dé la situación de que la cantidad de estudiantes inscriptos supere el cupo máximo.*

ANEXO B para la(s) carrera(s) XXX

(Un anexo distinto para cada carrera que tome la unidad curricular. En caso de que a dos o más carreras les corresponda información idéntica en este anexo, se utilizará el mismo anexo, explicitando cuáles son todas esas carreras.)

Esta(s) parte(s) del anexo incluye(n) los aspectos que son particulares de cada carrera que tome la unidad curricular.

B1) ÁREA DE FORMACIÓN

El área de formación (materia, según la anterior nomenclatura) identifica las grandes áreas temáticas ligadas a un sector de la ciencia o de la técnica. Cada comisión de carrera evaluará a qué área de formación corresponde la unidad curricular.

B2) UNIDADES CURRICULARES PREVIAS

Curso:

Examen:

(Las unidades curriculares previas serán definidas por las carreras que tomen la unidad curricular en cuestión, teniendo en cuenta los conocimientos exigidos que figuran en el programa.)