

NO FOLIO  
Comp. - 2000  
4

**FACULTAD DE CIENCIAS EXACTAS Y NATURALES**

1. DEPARTAMENTO: Computación
2. CUATRIMESTRE: Segundo de 2000
3. ASIGNATURA: Estadística y Data Mining.
4. CARRERA: Se ofrece para Licenciatura en Ciencias de la Computación , Licenciatura en Ciencias Biológicas. Licenciatura en Ciencias Químicas. Licenciatura en Ciencias Matemáticas.
5. CARACTER DE LA MATERIA: (Optativa)
6. NUMERO DE CODIGO DE CARRERA: 18
7. NUMERO DE CODIGO DE MATERIA: C .....
8. PUNTAJE: 3
9. PLAN DE ESTUDIOS AÑO: (1987 y 1993)
10. DURACION DE LA MATERIA: cuatrimestral
11. HORAS DE CLASE SEMANAL:
  - a) TEORICAS: 3 hs
  - b) LABORATORIO: 2 hs
  - c) PRACTICAS: 2 hs
  - d) SEMINARIOS: hs
12. CARGA HORARIA TOTAL: 7hs 105hs CUATRIMESTRALES
13. ASIGNATURAS CORRELATIVAS: Probabilidades y Estadística o Biometría o Estadística para químicos
14. FORMA DE EVALUACION: (Dos parciales y un trabajo de aplicación)
15. PROGRAMA Y BIBLIOGRAFIA: Adjuntas a esta hoja.

Fecha:

*AA*  
Dra. ANA MAEDO

*[Signature]*  
Dra. PATRICIA BORENSZTEJN  
DIRECTORA  
DEPTO. DE COMPUTACION  
F. C. E. y N. UBA

16. DOCENTES: Profesora Ana Silvia Haedo  
Ayudantes: Daniel Vazquez Vargas.  
Alejandra Christen-

## **Estadística y Data Mining**

### **Objetivo:**

Los metodos del analisis de Datos , tienen por objetivo proveer una síntesis global de un conjunto de datos. Basados en el conjunto total de datos no tienen hipótesis previas.

El principio que guía este curso es introducir a los estudiantes en técnicas estadísticas multivariadas. y familiarizarlos con el software que se utiliza en sus aplicaciones en la resolución de problemas diversos .

El dominio de los métodos estadísticos, conjuntamente con las herramientas informáticas de manejo y análisis de bases de datos de gran dimensión es un dominio de amplia difusión en la actualidad y de suma importancia en el desenvolvimiento profesional.

### **Programa:**

Análisis exploratorio y confirmatorio. Reseña histórica.

Técnicas descriptivas multidimensionales.

Representaciones gráficas multidimensionales.

Análisis en componentes principales.

Análisis factorial de correspondencias.

Métodos de clasificación y agrupamiento. Clasificación Jerárquica. K-medias. Caras de Chernov. Gráficos de estrellas. Gráficos de Rayos Sol. Gráficos de Andrews.

Relación entre Análisis factorial y Clasificación.

Árboles de decisión.

Métodos de segmentación.

Análisis discriminante.

Otros elementos de " Data mining" , redes neuronales .

### **Bibliografía:**

Benzecri J.P.& F. L'Analyse des données. Dunod. Paris.1980.

Berry Michael J. A., Linoff Gordon "Data Mining Techniques for Marketing Sales and Customer Support", John Wile& Sons, 1997.

Chatfield C., Collins A.J. Introduction to multivariate analysis. Chapman and Hall. London.1980.

Hoaglin D.C., Mosteller F., Tukey J.W., Exploring Data tables, Trends, and Shapes. Wiley. N.Y.1985.

Jambu M. Classification Automatique pour l'Analyse des données. Dunod. Paris.1978.

Johnson Dallas E. Métodos multivariados aplicados al análisis de datos. Thomson editores. Méjico.1998.

Johnson R.A. , Wichern Dean W. (1992) Applied Multivariate Statistical Analysis. 3rd De. Prentice Hall Inc. USA.

Lebart L., Morineau A., Tabard N. Techniques de la description statistique. Dunod. Paris.1977.

Dra. PATRICIA FORRESTIERI  
DIRECTORA  
DEPTO. DE COMPUTACION  
F. C. E. Y N. UBA

Lebart L., Morineau A., Piron Marie. Statistique exploratoire multidimensionnelle. Dunod. Paris.1995.

Pla , Laura E. (1986). Análisis. Multivariado: Método de Componentes Principales.

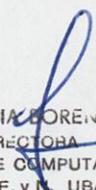
Monografía N° 27 Serie de matemática O.E.A.Washington.USA.

[MIC-SPI-TAY/94]. Machine Learning, Neural and Statistical Classification.

<http://ece.wpi.edu:8080/infoeng/textbook/node92.html>

[http://cns.bu.edu/pub/cn550/project97/algorithms/Decision\\_Trees.html](http://cns.bu.edu/pub/cn550/project97/algorithms/Decision_Trees.html)

Es conveniente la compra de software estadístico (Statistica, SPSS, SPAD. (para análisis de datos)) ya que su instalación en los laboratorios, permitiría no sólo el uso en el dictado de la materia, sino que posibilitaría dictado de curso de extensión , y asimismo desarrollar tareas de consultoría en análisis de datos de amplia demanda actual.

  
Dra. PATRICIA FORENSTEIN  
DIRECTORA  
DEPTO. DE COMPUTACION  
F. C. E. y N. UBA