

HORARIO

Todas las actividades académicas se desarrollarán en el Auditorio del Cimat

	lunes	martes	miércoles	jueves	viernes
09:00 a 9:30	Inscripciones				
09:30 a 10:00	G. Lugosi (C1)	G. Lugosi (C1)	G. Lugosi (C1)	G. Lugosi (C1)	G. Lugosi (C1)
10:00 a 10:30					
10:30 a 11:00					
11:00 a 11:30	C A F É				
11:30 a 12:00	Y. Lee (C2)	Y. Lee (C2)	S. Rosset (C3)	S. Rosset (C3)	S. Rosset (P5)
12:00 a 12:30					
12:30 a 13:00	V. Pérez-Abreu (P1)	C. Ludeña (P6)			
13:00 a 13:30					
13:30 a 14:00					
14:00 a 14:30					
14:30 a 15:00					
15:00 a 15:30					
15:30 a 16:00					
16:00 a 16:30	J. L. Marroquín (P2)	M. Nakamura (P4)		Y. Lee (C2)	
16:30 a 17:00					
17:00 a 17:30	M. Rivera (P3)	C. Ludeña (P7)			
17:30 a 18:00					
18:00 a 18:30				Presentación postgrado	

19:30	Taquiza en el Cimatel
-------	-----------------------

RESUMENES

C1:

Aprendizaje estadístico

Gábor Lugosi (Universidad Pompeu Fabra)

En este curso se repasarán algunos conceptos de aprendizaje estadístico. Se investigarán los métodos de la minimización empírica del riesgo, la minimización estructural de riesgo, y algunos algoritmos de minimización de funciones convexas de riesgo, como "boosting". Entre las herramientas matemáticas esenciales se discutirán algunas desigualdades de concentración, las nociones de la dimensión Vapnik-Chervonenkis, promedios de Rademacher, etc.

En la segunda parte del curso se discutirán algunas ideas básicas de la predicción on-line de sucesiones "individuales". Este modelo de aprendizaje está estrechamente relacionado con modelos de aprendizaje en la teoría de juegos. Se presentarán algunos resultados sobre algoritmos aleatorizados de predicción.

C2:

Kernel methods in a regularization framework

Yoonkyung Lee (Ohio State University)

Kernel methods for statistical learning are introduced with smoothing splines and support vector machines as primary examples. The commonality between them as methods of regularization is to be emphasized. For theoretical underpinnings of the methods, general treatment of reproducing kernels will be given with discussion of regularization in reproducing kernel Hilbert spaces and the representer theorem. In addition, it is illustrated how feature selection for kernel methods can be dealt with in a principled way using the structured kernels associated with functional ANOVA decomposition for enhanced interpretability.

C3:

Regularization and boosting

Saharon Rosset (IBM)

*I will start with an introduction to practical Statistical Learning, with some examples of applications, discussion of some fundamental methods and their statistical properties and demonstration of their performance. The first part of this short course will follow the early chapters of *The Elements of Statistical Learning* by Hastie, Tibshirani and Friedman. I will then concentrate on regularization in Statistical Learning, discussing its sta-*

tistical role in shrinkage and model selection, and computational aspects related to derivation of regularization paths. I will introduce Boosting, and discuss its interpretation as approximate L_1 regularization in high dimensional spaces. If time permits, I will also present my recent work on exact L_1 regularization in very high -- even infinite -- feature spaces.

P1:

Desigualdades e identidades de correlación

Victor Pérez-Abreu (CIMAT)

En la primera parte de la conferencia hablaremos de la importancia del trabajo de Jacobo Bernoulli en probabilidad y posteriormente presentaremos algunos resultados y aplicaciones simples de desigualdades de correlación.

P2:

Métodos Bayesianos para procesamiento de imágenes

José Luis Marroquín (CIMAT)

Se presentará el uso de métodos de Estimación Bayesiana, en particular, con modelos a priori de Campos Aleatorios Markovianos, para la solución de algunos problemas de procesamiento de imágenes y visión computacional, haciendo especial énfasis en el diseño de algoritmos computacionalmente eficientes. Se presentarán soluciones a problemas de segmentación de imágenes, con aplicaciones a la segmentación de imágenes médicas y se indicará cómo pueden extenderse estos métodos para la solución de otros problemas, como: la estimación de la profundidad a partir de imágenes estereoscópicas; la estimación del movimiento a partir de secuencias de imágenes, etc.

P3:

Quadratic Markovian probability fields for image binary segmentation

Mariano Rivera (CIMAT)

A Markov Random Field model for image binary segmentation computes the probability that each pixel belongs to a given class. We show that if a real valued field is computed instead of a binary one as in graph cuts based methods, then the resultant cost function has noticeable computational and performance advantages. The performance of the method is demonstrated for the interactive binary segmentation task (segmentation by trimaps) and compared with the binary counterpart.

P4:

**Algunas reflexiones sobre tipología de modelos
con motivo de nichos ecológicos**

Miguel Nakamura (CIMAT)

Se utilizará como plataforma el problema biológico de predicción de nichos de especies —el cual es en principio un problema de clasificación binaria— con el objetivo de motivar consideraciones generales en torno a alcances y características de clases de técnicas. En particular, se discutirá el significado que puede tener el concepto de “modelo estadístico” para distintos gremios.

P5:

Data analytics for marketing decision support

Saharon Rosset (IBM)

I will give an overview of the issues in applying data mining and analytics tools to marketing decision support, primarily for marketing and sales optimization purposes. In particular, I will:

- review some challenges involved in designing analytics tools for real-life marketing decision support;

- discuss different analytical approaches to addressing these problems in a practically feasible manner;

- present a detailed case study based on my experience in designing and implementing a system for customer wallet estimation --- an important business problem for which no "standard" solutions exist.

P6:

Adaptive estimation of linear functionals by model selection

Carenne LUDEÑA

Instituto Venezolano de Investigaciones Científicas, Venezuela
e-mail: cludena@ivic.ve

Abstract

Consider the following model :

$$Y(t) = \langle s, t \rangle + \frac{1}{\sqrt{n}}L(t), \text{ for all } t \in H,$$

where H is an Hilbert space with inner product \langle, \rangle and L is a centered Gaussian process over H such that for all $t, u \in H$, $\text{Cov}(L(t), L(u)) = \langle t, u \rangle$.

Model selection (MS) refers to the problem of selecting a model m leading to a least squares estimator \hat{s}_m with close to minimal risk value among a certain collection of possible models $S_m \subset H$. For each m the quadratic risk is

$$E\|s - \hat{s}_m\|^2 = \|s_m - s\|^2 + d_{m,n}$$

where $d_{m,n}$ is a variance term independent of s , and s_m is the projection of s over S_m .

Thus the "correct" choice of model m follows by estimating the bias term in the above bound, which can be done easily, and minimizing the sum.

If, however, we consider the problem of estimating $T(s)$, where T is some linear functional, the bias term cannot be simplified nor estimated directly and a new model selection procedure must be developed.

We show the procedure is adaptive, up to a constant, and give non asymptotic inequalities for the quadratic risk. An application using a MRA for the problem of estimating a signal or its r^{th} derivative at a given point is developed and minimax rates are seen to hold uniformly over Besov spaces. Simulations are included to illustrate the procedure's performance.

A statistical view at ill posed inverse problems and regularization methods

Carenne LUDEÑA

Instituto Venezolano de Investigaciones Científicas, Venezuela

e-mail: cludena@ivic.ve

We are interested in recovering an unobservable signal x_0 based on observations

$$y(t_i) = A(x_0)(t_i) + \varepsilon_i,$$

where $A : X \rightarrow Y$ is a compact operator with unbounded inverse, X, Y are Hilbert spaces and $t_i, i = 1, \dots, n$ is a fixed observation scheme. Although $A(x_0) \in Y$, we cannot observe this function directly and inference on the unknown function $x_0 \in X$ will depend on the observed data $y(t_i), i = 1, \dots, n$.

The ill-posedness of the problem requires the introduction of some kind of regularization procedure, indexed by some regularization parameter α that measures the closeness of the regularized and the original (unregularized) inverse problem. Rules (and algorithms) for the choice of these regularization parameters, with or without prior information on the smoothness properties of the solution are a central issue. Our main interest will be the adaptive choice of α . For this it is necessary to introduce an appropriate penalty function. As examples of regularization procedures we consider linear model selection, Tikhonov methods and iterative procedures such as steepest descent methods.

Based on concentration inequalities techniques we derive non asymptotic optimal upper bounds of the mean square error for each regularization method.



Programa

- Lunes 22

- 9:15– 9:30 Inscripción
- 9:30–11:00 Aprendizaje estadístico
Gabor Lugosi
- 11:00–11:30 Café y galletas
- 11:30–13:00 Kernel methods in a regularization framework
Yoonkyung Lee
- 13:00–14:00 Desigualdades e identidades de correlación
Victor Pérez–Abreu
- 14:00–16:00 Comida
- 16:00–17:15 Métodos Bayesianos para procesamiento de imágenes
José Luis Marroquín
- 17:15–18:15 Quadratic Markovian probability fields for image binary segmentation. **Mariano J. J. Rivera**
- 19:30–22:00 Taquiza en el CIMATEL

- Martes 23

- 9:30–11:00 Aprendizaje estadístico
Gabor Lugosi
- 11:00–11:30 Café y galletas

- 11:30–13:00 Kernel methods in a regularization framework
Yoonkyung Lee
- 13:00–14:00 Adaptive estimation of linear functionals by model selection
Carenne Ludeña
- 14:00–16:00 Comida
- 16:00–17:15 Algunas reflexiones sobre tipología de modelos con motivo de nichos ecológicos. **Miguel Nakamura**
- 17:15–18:30 A statistical view of ill posed inverse problems and regularization methods. **Carenne Ludeña**

- Miércoles 24

- 9:30–11:00 Aprendizaje estadístico
Gabor Lugosi
- 11:00–11:30 Café y galletas
- 11:30–14:00 Regularization and boosting
Saharon Rosset
- 14:00–16:00 Comida
- 16:00– Tarde libre.

- Jueves 25

- 9:30–11:00 Aprendizaje estadístico
Gabor Lugosi
- 11:00–11:30 Café y galletas
- 11:30–14:00 Regularization and boosting
Saharon Rosset
- 14:00–16:00 Comida
- 16:00–17:30 Kernel methods in a regularization framework
Yoonkyung Lee
- 17:30–18:00 Estudios de postgrado en Probabilidad y Estadística en CIMAT. **Eloísa Díaz–Francés**

- Viernes 26

- 9:30–11:00 Aprendizaje estadístico

- Gabor Lugosi**

- 11:00–11:30 Café y galletas

- 11:30–14:00 Data analytics for marketing decision support

- Saharon Rosset**

- 14:00–14:10 Clausura.