

Formulario de Aprobación Curso de Posgrado

Asignatura: A short course in model selection and sparse models.

(Si el nombre contiene siglas deberán ser aclaradas)

Profesor de la asignatura ¹:

Dr. Jean-Marc Azaïs, Professeur des Universités, Institut de Mathématiques de Toulouse
(título, nombre, grado o cargo, Instituto o Institución)

Profesor Responsable Local ¹:

Dr. Ernesto Mordecki, Profesor Titular (Grado 5), Centro de Matemática, Facultad de Ciencias
(título, nombre, grado, Instituto)

Otros docentes de la Facultad:

(título, nombre, grado, Instituto)

Docentes fuera de Facultad:

Dr. Fabrice Gamboa, Professeur des Universités, Institut de Mathématiques de Toulouse
Dr. Guillaume Obozinski, Chargé de recherche, WILLOW, École Normale Supérieure de Paris
(título, nombre, cargo, Institución, país)

Instituto ó Unidad:

Departamento ó Area: Instituto Franco-Uruguayo de Matemática (IFUM) y Centro de Matemática

¹ Agregar CV si el curso se dicta por primera vez.

(Si el profesor de la asignatura no es docente de la Facultad se deberá designar un responsable local)

Fecha de inicio y finalización: 28 de febrero al 11 de marzo de 2011

Horario y Salón: horario a determinar. Lugar: Centro de Matemática, Facultad de Ciencias, Universidad de la República, Montevideo.

Horas Presenciales: 35

(sumar horas directas de clase – teóricas, prácticas y laboratorio – horas de estudio asistido y de evaluación)
Se deberán discriminar las mismas en el ítem Metodología de enseñanza.

Nº de Créditos: 6

Público objetivo y Cupos: Estudiantes de la Maestría en Ingeniería Matemática, otros estudiantes de posgrado e investigadores de la región interesados en los tópicos de selección de modelos y modelos esparsos.

(si corresponde, se indicará el número de plazas, mínimo y máximo y los criterios de selección. Si no existe indicación particular para el cupo máximo, el criterio general será el orden de inscripción en el Depto. de Posgrado, hasta completar el cupo asignado)

Objetivos: Model selection is an important issue of modern statistics. Besides the classical results using penalties based on the size of the model: C_p AIC BIC, a new kind of techniques have appeared with L^1 penalties and the LASSO (Tibshirani, 1996) and the work of Donoho and Candès on sparse models.

Conocimientos previos exigidos: Probabilidad e inferencia estadística

Conocimientos previos recomendados: Modelos estadísticos, aprendizaje automático

Metodología de enseñanza: El curso será dictado en los idiomas Inglés y Español. Los métodos y técnicas introducidas serán presentados tanto desde el punto de vista teórico como de la implementación práctica mediante recursos informáticos.

Se dictarán 10 clases de tres horas cada una.

Habrà 5 horas de consulta durante el período del curso.
Se estima una cantidad de 55 horas en las que el estudiante trabaje de forma independiente para asimilar el contenido del curso y preparar el examen final.

(comprende una descripción de las horas de clase asignadas y su distribución en horas de práctico, horas de teórico, horas de laboratorio, etc. si corresponde)

Forma de evaluación: Examen oral final en idioma español.

Temario:

Mainly we will address new trends in mathematical statistics around sparse models. This includes recent development on L^1 methods. In particular, the following topics will be addressed:

- The problem of over-fitting when the sizes increase ([11]),
- Selection of models by classical tools: Mallows C_p , BIC, penalized likelihood,..
- Model selection using Lasso, ([12], [10], [9]),
- Dantzig selector and compressed sensing ([2],[3],[1],[4],[8],[5],[7],[6]),
- Oracle inequalities,
- Lasso and machine learning, application to genetics.

Bibliografía:

(título del libro-nombre del autor-editorial-ISBN-fecha de edición)

- [1] Jian-Feng Cai, Emmanuel J. Candès, and Zuowei Shen. A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.*, 20(4):1956-1982, 2010.
- [2] Emmanuel Candès and Terence Tao. The Dantzig selector: statistical estimation when p is much larger than n . *Ann. Statist.*, 35(6):2313-2351, 2007.
- [3] Emmanuel Candès and Terence Tao. Rejoinder: "The Dantzig selector: statistical estimation when p is much larger than n " [Ann. Statist. 35 (2007), no. 6, 2313-2351; mr2382644]. *Ann. Statist.*, 35(6):2392-2404, 2007.
- [4] Emmanuel J. Candès. The restricted isometry property and its implications for compressed sensing. *C. R. Math. Acad. Sci. Paris*, 346(9-10):589-592, 2008.
- [5] David Donoho and Jiashun Jin. Asymptotic minimaxity of false discovery rate thresholding for sparse exponential data. *Ann. Statist.*, 34(6):2980-3018, 2006.
- [6] David L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289-1306, 2006.
- [7] David L. Donoho. For most large underdetermined systems of equations, the minimal l_1 -norm near-solution approximates the sparsest near-solution. *Comm. Pure Appl. Math.*, 59(7):907-934, 2006.
- [8] David L. Donoho and Yaakov Tsaig. Fast solution of l_1 -norm minimization problems when the solution may be sparse. *IEEE Trans. Inform. Theory*, 54(11):4789-4812, 2008.
- [9] Lukas Meier, Sara van de Geer, and Peter Bühlmann. The group Lasso for logistic regression. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 70(1):53-71, 2008.
- [10] Lukas Meier, Sara van de Geer, and Peter Bühlmann. High-dimensional additive modeling. *Ann. Statist.*, 37(6B):3779-3821, 2009.
- [11] Ryuei Nishii. Asymptotic properties of criteria for selection of variables in multiple regression. *Ann. Statist.*, 12(2):758-765, 1984.
- [12] Sara A. van de Geer and Peter Bühlmann. On the conditions used to prove oracle results for the Lasso. *Electron. J. Stat.*, 3:1360-1392, 2009.