

**Formulario de aprobación de curso de posgrado/educación permanente**

**Asignatura: “Tópicos en CART y en Random Forest”**

(Si el nombre contiene siglas deberán ser aclaradas)

**Modalidad:**

(posgrado, educación permanente o ambas)

**Posgrado**

**X**

**Educación permanente**

**Profesor de la asignatura <sup>1</sup>: Prof. Jean-Michel Poggi, Professeur des Universités, (Univ Paris Descartes & LMO, Univ Paris-Sud, Orsay).**

(título, nombre, grado o cargo, instituto o institución)

**Profesor Responsable Local <sup>1</sup>: Dr. Mathias Bourel, Grado 3 DT, IMERL.**

(título, nombre, grado, instituto)

**Otros docentes de la Facultad:**

(título, nombre, grado, instituto)

**Docentes fuera de Facultad:**

(título, nombre, cargo, institución, país)

<sup>1</sup> 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)

[Si es curso de posgrado]

**Programa(s) de posgrado: Maestría en Ingeniería Matemática**

**Instituto o unidad: Instituto de Matemática y Estadística “Prof. Ing. Rafael Laguardia”.**

**Departamento o área:**

**Horas Presenciales: 20**

(se deberán discriminar las horas en el ítem Metodología de enseñanza)

**Nº de Créditos: 4**

[Exclusivamente para curso de posgrado]

(de acuerdo a la definición de la UdelaR, un crédito equivale a 15 horas de dedicación del estudiante según se detalla en el ítem Metodología de enseñanza)

**Público objetivo:** Estudiantes de la Maestría en Ingeniería Matemática, Maestría en Matemática, Ingenieros y otros estudiantes interesados.

**Cupos:** Sin cupo

**Objetivos:** Introducir los aspectos metodológicos de CART & Random Forest así como algunas contribuciones recientes en cuanto a la aplicación de los árboles a la estadística espacial y selección de variables. Aplicar dichas técnicas a conjuntos de datos reales e interpretar los resultados obtenidos. Acercar el estudiante al empleo de los paquetes para el análisis estadístico de datos disponibles en el ambiente de desarrollo de software libre R (<http://www.r-project.org/>).

**Conocimientos previos exigidos:** introducción a la probabilidad y estadística, álgebra lineal, cálculo matricial.

**Conocimientos previos recomendados:** un curso de R

---

**Metodología de enseñanza:**

Descripción de la metodología: Teórico-Práctico

Detalle de horas:

- Horas de clase (teórico): 12
- Horas de clase (práctico): 4
- Horas de clase (laboratorio):
- Horas de consulta: 8
- Horas de evaluación:
  - Subtotal de horas presenciales: 24
- Horas de estudio: 18
- Horas de resolución de ejercicios/prácticos: 18
  - Total de horas de dedicación del estudiante: 60

---

**Forma de evaluación:** Entrega de ejercicios

---

**Temario:**

- **1. Árboles de clasificación y de regresión (CART).** Introducción. Partición de nodos, construcción del árbol maximal y algoritmo de poda. Selección de variables con CART. Extensiones y generalizaciones. El paquete rpart en R.
- **2. CART para datos espaciales.** Nuevo criterio de partición de nodos para datos espaciales en contexto binario.
- **3. Random Forests, Importancia de variable y selección de variables.** Bagging y Random Forests (RF). Extensiones y variantes, algunos resultados teóricos. Error out-of-bag y selección de variables con RF. Los paquetes randomForest y VSURF en R.
- **4. Aplicación:** Una aproximación basada en RF para funcionales fisiológicos. Selección de variables: hacia una clasificación del nivel de estrés de conductores. Introducción y motivación. Variables del funcional fisiológica y onduladas. Medida de la importancia de bloques de variables. Selección de variables funcionales usando RF. Clasificación del nivel de estrés de conductores. RF para Big data: algunos problemas y soluciones para proponer un versión escalable de RF.

**Bibliografía:**

- Breiman, L., Friedman, J., Olshen, R. et Stone, C. [1984]. Classification and Regression Trees . Chapman & Hall, New York.
- Bar-Hen, A., Gey, S. et Poggi, J.-M. [2015]. Influence measures for CART classification trees. Journal of Classification, 32 (1), 21–45.
- Therneau, T., Atkinson, B. et Ripley, B. [2015]. rpart : Recursive Partitioning and Regression Trees
- Bar-Hen, A., Gey, S. et Poggi, J.-M. [2015]. Influence measures for CART classification trees. Journal of Classification, 32 (1), 21–45.
- Bar-Hen, A., Gey, S. et Poggi, J.-M. [2018]. Spatial CART Classification Trees. 2018. <hal-01837065>
- A. Bar-Hen, J.-M. Poggi [2016] Influence Measures and Stability for Graphical Models. Journal of Multivariate Analysis, Vol. 47, 145-154
- Breiman, L. [2001]. Random forests. Machine learning , 45 (1), 5–32.
- Genuer, R., Poggi, J.-M. et Tuleau-Malot, C. [2010]. Variable selection using random forests. Pattern Recognition Letters , 31 (14), 2225–2236.
- Genuer, R., Poggi, J.-M. et Tuleau-Malot, C. [2015]. Vsurf : An R package for variable selection using random forests. The R Journal , 7 (2), 19–33.
- Gregorutti, B., Michel, B. et Saint-Pierre, P. [2015]. Grouped variable importance with random forests and application to multiple functional data analysis. Computational Statistics & Data Analysis, 90, 15–35.
- El Haouij N., Poggi J.-M., Ghozi R., Sevestre Ghalila S., Jaïdane M [2018], Random Forest-Based Approach for Physiological Functional Variable Selection: Towards Driver's Stress Level Classification, Statistical Methods & Applications, 1-29, 2018
- M. Bobbia, F.-X. Jollois, J.-M. Poggi, B. Portier [2011] Quantifying local and background contributions to PM10 concentrations in Haute-Normandie, using random forests, Environmetrics, 22, 758-768
- Genuer, R., Poggi, J.-M., Tuleau-Malot, C. et Villa-Vialaneix, N. [2017]. Random Forests for Big Data, Big Data Research, 9, 2017, 28-46

**Datos del curso**

---

**Fecha de inicio y finalización: 10/2/2020 - 17/2/2020**

**Horario y Salón: IMERL.**

**Arancel:**

[Si la modalidad no corresponde indique "no corresponde". Si el curso contempla otorgar becas, indíquelo]

**Arancel para estudiantes inscriptos en la modalidad posgrado:**

**Arancel para estudiantes inscriptos en la modalidad educación permanente:**

---