

Formulario de Aprobación Curso de Posgrado

Asignatura: Seminario de Ciencia de Datos

(Si el nombre contiene siglas deberán ser aclaradas)

Profesor de la asignatura ¹: Dr. Emilio Carrizosa, Universidad de Sevilla, España.

(título, nombre, grado o cargo, Instituto o Institución)

Profesor Responsable Local ¹: Dr. Héctor Cancela, Gr. 5, INCO

(título, nombre, grado, Instituto)

Otros docentes de la Facultad:

(título, nombre, grado, Instituto)

Docentes fuera de Facultad:

(título, nombre, cargo, Institución, país)

Instituto ó Unidad: Instituto de Computacion

Departamento ó Area: Departamento de Investigación Operativa

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)

Horas Presenciales: 4,5 hs

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

Nº de Créditos: 3

(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 la enseñanza)

Público objetivo y Cupos: sin cupos

(si corresponde, se indicará el número de plazas, mínimo y máximo y los criterios de selección. Asimismo, se adjuntará en nota aparte los fundamentos de los cupos propuestos. 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)

El curso esta dirigido a estudiantes de: cualquier programa de posgrado en ramas de la ingeniería o disciplinas afines.

Objetivos:

El propósito del curso es el presentar y discutir algunos temas fundamentales en el área de Ciencia de Datos (Data Science), focalizando particularmente en el papel que tienen las herramientas de modelado y los modelos de Programación Matemática.

Conocimientos previos exigidos: -

Conocimientos previos recomendados: -

Metodología de enseñanza:

(comprende una descripción de la metodología de enseñanza y de las horas dedicadas por el estudiante a la asignatura, distribuidas en horas presenciales -de clase práctica, teórico, laboratorio, consulta, etc.- y no presenciales de trabajo personal del estudiante)

La primer parte del curso comprende tres clases presenciales expositivas, en las cuales se presentarán los diferentes conceptos y las herramientas de modelado, particularmente de programación matemática, que pueden emplearse en este contexto. En las

mismas se presentará también un conjunto de artículos científicos recientes que profundizan en los distintos aspectos presentados.

En la segunda parte del curso, cada estudiante deberá elegir uno de estos artículos, para estudiarlo en profundidad (para lo que deberá también apoyarse en el conjunto de la bibliografía del curso) y elaborar un resumen que demuestre haber comprendido las características de la técnica expuesta en el artículo, su campo de aplicación, y el desempeño esperable de la misma. Se dará apoyo de consultas a través de correo o foros.

- Horas clase (teórico): 4,5 hs
- Horas clase (práctico): 0
- Horas clase (laboratorio): 0
- Horas consulta: 0
- Horas evaluación: 0
 - Subtotal horas presenciales: 4.5
- Horas estudio: 12
- Horas resolución ejercicios/prácticos:
- Horas proyecto final/monografía: 30
 - Total de horas de dedicación del estudiante: 46

Forma de evaluación:

Para la evaluación se tendrá en cuenta:

Asistencia a clase.

Lectura de un artículo vinculado a los tópicos abordados en el curso y redacción de un resumen del mismo.

Temario:

Block 1: Dimensionality reduction in Data Science: seeing what you get.

Discussion on Principal Component Analysis and Multidimensional Scaling

Block 2: An introduction to Classification and Regression methods, with special emphasis in Support Vector Machines and Classification and Regression Trees

Block 3: Sparse methods: straight to the point.

Sparse regression, sparse PCA, sparse SVM

Bibliografía:

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

E Carrizosa, V Guerrero. rs-Sparse principal component analysis: A mixed integer nonlinear programming approach with VNS. *Computers & Operations Research* 52, 349-354, 2014

E Carrizosa, V Guerrero, D Romero Morales. Visualizing proportions and dissimilarities by space-filling maps: a large neighborhood search approach. *Computers & Operations Research* 78, 369-380, 2017

E Carrizosa, A Nogales-Gómez, D Romero Morales. Clustering categories in support vector machines. *Omega* 66, 28-37, 2017

E Carrizosa, AV Olivares-Nadal, P Ramírez-Cobo. A sparsity-controlled vector autoregressive model. *Biostatistics* 18 (2), 244-259, 2016



Facultad de Ingeniería Comisión Académica de Posgrado

E Carrizosa, D Romero Morales. Supervised classification and mathematical optimization. Computers & Operations Research 40 (1), 150-165, 2013

Otros artículos y bibliografía a ser difundida en clase.



Facultad de Ingeniería Comisión Académica de Posgrado

Datos del curso

Fecha de inicio y finalización: setiembre 2017

Horario y Salón: teóricos: miércoles 27, jueves 28, viernes 29 setiembre, 17 a 18:30hs, salón a confirmar.
