

62236 - Análisis avanzado de datos

Información del Plan Docente

Año académico: 2024/25

Asignatura: 62236 - Análisis avanzado de datos

Centro académico: 110 - Escuela de Ingeniería y Arquitectura

Titulación: 534 - Máster Universitario en Ingeniería Informática

Créditos: 3.0

Curso: 2

Periodo de impartición: Primer semestre

Clase de asignatura: Optativa

Materia:

1. Información básica de la asignatura

La asignatura tiene como objetivo la formación en la construcción de modelos estocásticos que expliquen las relaciones entre variables o individuos en estudios observacionales. Esto requiere conocer herramientas en la estimación máximo-verosímil y bayesiana, algoritmo EM, métodos MCMC y de remuestreo.

Se introducen técnicas estadísticas y se aprende a seleccionar la más adecuada para obtener soluciones en el análisis de grandes bases de datos. Mediante técnicas de regresión se calculan predicciones de la respuesta y cotas de su error en situaciones de aprendizaje supervisado. Para situaciones en que no existe variable respuesta, sino un conjunto de variables que representan la realidad, los métodos de aprendizaje no supervisado reconocen patrones de variables y casos, con el objetivo de caracterizarlos y de reducir la dimensión.

2. Resultados de aprendizaje

1. Interpretar datos observacionales o experimentales, extraer la información que contienen, identificar las relaciones entre ellos y evaluar hipótesis en presencia de incertidumbre y variabilidad, interpretando adecuadamente sus resultados.
2. Comprender los métodos de estimación, por máxima-verosimilitud y bayesianos, conocer las herramientas y algoritmos para la estimación en grandes bases de datos.
3. Aplicar procedimientos estadísticos de construcción y validación de modelos empíricos que expresan la relación entre una variable respuesta y otras variables cuyo valor se puede conocer.
4. Utilizar las técnicas más relevantes de análisis multivariante que contribuyen a explicar las relaciones entre los datos e identificar patrones cuando no hay una variable respuesta.
5. Aplicar técnicas de aprendizaje, minería de datos y minería de procesos para la extracción de conocimiento en entornos que manejen grandes cantidades de datos a través de la Web.
6. Aplicar técnicas matemáticas para el análisis de grandes cantidades de datos en la Web.

3. Programa de la asignatura

- Introducción
 - Análisis exploratorio de datos y aprendizaje estadístico.
 - Muestreo e inferencia estadística: estimación puntual y por intervalo, contrastes de hipótesis.
 - Verosimilitud: Estimación por máxima-verosimilitud, test de cociente de verosimilitudes.
 - Teoría estadística de la decisión. Métodos bayesianos.
 - Algoritmo EM y método MCMC.
 - Simulación estadística.
- Reconocimiento de relaciones explícitas: Modelos de regresión
 - Modelo lineal general, covariables y factores, análisis de la varianza.
 - Procedimientos automáticos de construcción de modelos: best subset, stepwise.
 - Validación, validación cruzada, métodos bootstrap.
 - Modelos con respuesta no gaussiana: GLM y GAM.
- Reconocimiento de patrones asistido
 - Modelos de regresión logística binaria y multinomial.
 - Modelos de regresión logística multinomial.
 - Modelos loglineales.
- Reconocimiento de patrones no supervisado

- Análisis de conglomerados, k-medias.
- Clúster jerárquico.

4. Actividades académicas

La asignatura consta de 3 créditos ECTS que suponen una dedicación estimada por parte del alumno de 75 horas (35 horas de docencia con el profesor y 40 horas no presenciales) distribuidas del siguiente modo:

- 30 horas, aproximadamente, de actividades formativas con el profesor (clases magistrales incluyendo seminarios profesionales, resolución de problemas y casos, y prácticas de laboratorio).
- 20 horas de trabajo en grupo.
- 20 horas de trabajo y estudio individual efectivo.
- 5 horas dedicadas a distintas pruebas de evaluación.

5. Sistema de evaluación

El estudiante debe demostrar que ha alcanzado los resultados de aprendizaje previstos mediante las siguientes actividades de evaluación:

Seguimiento del trabajo realizado en las sesiones prácticas. Trabajo desarrollado en las sesiones prácticas relativo al análisis de datos [20%]. Resultados de aprendizaje: 1, 2, 3 y 4.

Trabajo dirigido. Un proyecto individual o en grupo en el que se podrán en práctica los conocimientos y habilidades adquiridos en la asignatura. En la evaluación del trabajo tutorado propuesto a lo largo del cuatrimestre se tendrá en cuenta tanto la memoria presentada, como la idoneidad y originalidad de la solución propuesta. [70%]. Resultados de aprendizaje: 1 a 6.

Presentación de forma oral y debate sobre el desarrollo y resultados del trabajo académico (10%). Resultados de aprendizaje: 1 a 6.

El estudiante que no opte por el procedimiento de evaluación descrito anteriormente, no supere dichas pruebas durante el periodo docente o que quisiera mejorar su calificación, tendrá derecho a realizar una prueba global que será programada dentro del periodo de exámenes correspondiente a la primera o segunda convocatoria.

6. Objetivos de Desarrollo Sostenible

7 - Energía Asequible y No Contaminante
15 - Vida de Ecosistemas Terrestres