



## Guía docente

# 230338 - APATP - Aprendizaje Automático: de la Teoría a la Práctica

Última modificación: 17/06/2024

**Unidad responsable:** Escuela Técnica Superior de Ingeniería de Telecomunicación de Barcelona

**Unidad que imparte:** 739 - TSC - Departamento de Teoría de la Señal y Comunicaciones.

**Titulación:** GRADO EN INGENIERÍA DE TECNOLOGÍAS Y SERVICIOS DE TELECOMUNICACIÓN (Plan 2015). (Asignatura optativa).

GRADO EN INGENIERÍA ELECTRÓNICA DE TELECOMUNICACIÓN (Plan 2018). (Asignatura optativa).

**Curso:** 2024

**Créditos ECTS:** 2.0

**Idiomas:** Catalán, Castellano

## PROFESORADO

**Profesorado responsable:** Josep Vidal

**Otros:** Josep Vidal, Verónica Vilaplana

## CAPACIDADES PREVIAS

Estadística

## REQUISITOS

GRADO EN INGENIERÍA DE TECNOLOGÍAS Y SERVICIOS DE TELECOMUNICACIÓN (GRETST)

PROBABILIDAD Y ESTADÍSTICA - Prerrequisito

GRADO EN INGENIERÍA ELECTRÓNICA DE TELECOMUNICACIÓN (GREELEC)

PROBABILIDAD Y PROCESOS ESTOCÁSTICOS - Prerrequisito

## METODOLOGÍAS DOCENTES

El tiempo se repartirá entre clases de teoría y laboratorio en Python.

## OBJETIVOS DE APRENDIZAJE DE LA ASIGNATURA

Esta asignatura realiza una revisión de la teoría de la clasificación y la regresión basados en datos desde una perspectiva matemática y al mismo tiempo aplicada a diferentes áreas. Se pretende dar una visión general de la teoría de decisión bayesiana, técnicas de clasificación paramétrica (regresión logística, redes neuronales, árboles de decisión, combinación de clasificadores, etc) y aprendizaje no supervisado con una experimentación simultánea en diferentes aplicaciones de entre las que destacamos decisión sobre datos biomédicos, clasificación de imágenes, detección de eventos, detección de spam, etc...

Los créditos de la asignatura se reparten en créditos teóricos y créditos prácticos (laboratorio en Python) para paralelizar los conocimientos teóricos con el desarrollo de aplicaciones.

En las últimas semanas del curso, todos los alumnos participarán en una competición interna de machine learning propuesta por los profesores.



## HORAS TOTALES DE DEDICACIÓN DEL ESTUDIANTADO

Tipo	Horas	Porcentaje
Horas grupo grande	10,0	20.00
Horas aprendizaje autónomo	30,0	60.00
Horas grupo pequeño	10,0	20.00

**Dedicación total:** 50 h

## CONTENIDOS

### 1. Introducción al aprendizaje automático

**Descripción:**

Introducción al aprendizaje automático. Aprendizaje supervisado y no supervisado. Regresión y clasificación

**Actividades vinculadas:**

Práctica de laboratorio 1. Introducción a Python, Google Colab y Scikit-Learn. Lectura de marcos de datos y análisis de datos exploratorios utilizando el conjunto de datos Iris y el conjunto de datos de Cáncer de mama. Ejemplos simples que muestran cómo resolver tareas supervisadas (clasificación) y no supervisadas (reducción de dimensionalidad y agrupamiento) utilizando el conjunto de datos Iris

**Dedicación:** 4h 15m

Grupo grande/Teoría: 2h

Actividades dirigidas: 2h

Aprendizaje autónomo: 0h 15m

### 2. Regresión

**Descripción:**

Regresión. Modelos, funciones de costos y entrenamiento. Regresión lineal y polinómica. Entrenamiento de descenso en gradiente. Sobreajuste, desajuste y generalización. Regularización: lazo, cresta y red elástica. Validación. Medidas de desempeño: MSE, MAP, coeficiente R, correlación de Pearson. Desafíos en el aprendizaje asociados a la calidad de los datos.

**Actividades vinculadas:**

Laboratorio 2. Regresión lineal y regularización utilizando el conjunto de datos MTCars. Solución de forma cerrada y solución mediante descenso de gradiente. Regresión polinomial.

**Dedicación:** 4h 15m

Grupo grande/Teoría: 2h

Actividades dirigidas: 2h

Aprendizaje autónomo: 0h 15m

### 3. Clasificación

**Descripción:**

Clasificación de dos clases, múltiples clases y múltiples etiquetas. Regresión logística. Límites de decisión. Función de costo y entrenamiento de gradiente. Medidas de desempeño: exactitud, precisión, recuperación, puntuación F, matriz de confusión, ROC. Regularización.

**Actividades vinculadas:**

Laboratorio 3. Regresión logística utilizando el conjunto de datos Iris y el conjunto de datos Cáncer de mama. Matriz de confusión y métricas de clasificación. Divisiones de prueba de tren. Presentación del concurso de aprendizaje automático.

**Dedicación:** 4h 15m

Grupo grande/Teoría: 2h

Actividades dirigidas: 2h

Aprendizaje autónomo: 0h 15m

### 4. Redes neuronales

**Descripción:**

Redes neuronales. Estructuras. Regresión/clasificación. Funciones de costos. Algoritmo de retropropagación. Técnicas de descenso de gradientes. Entrenamiento por lotes. Sobreajuste y generalización. Regularización y evitación del sobreajuste. RNN y DNN. Reglas de una competencia de aprendizaje automático. Validación de plegado K.

**Actividades vinculadas:**

Laboratorio 4. Clasificador de redes neuronales multicapa. Tuberías de Sklearn. Búsqueda de hiperparámetros mediante GridSearchCV. Regularización. Aplicación a conjuntos de datos Iris y MNIST.

**Dedicación:** 4h 15m

Grupo pequeño/Laboratorio: 2h

Actividades dirigidas: 2h

Aprendizaje autónomo: 0h 15m

### 5. Árboles de decisión

**Descripción:**

Árboles de decisión. Entrenamiento, sobreajuste y poda. Desempeño en términos de sesgo y varianza. Bosque aleatorio y aprendizaje ensamble.

**Actividades vinculadas:**

Laboratorio 5. Árboles de decisión: entrenamiento, visualización, regularización. Clasificadores de conjunto: bosque aleatorio, embolsado, impulso. Clasificadores de votación (blandos y duros). Aplicación a conjuntos de datos Iris y MNIST.

**Dedicación:** 4h 15m

Grupo grande/Teoría: 2h

Actividades dirigidas: 2h

Aprendizaje autónomo: 0h 15m

### 6. Aprendizaje no supervisado

**Descripción:**

Aprendizaje automático no supervisado. Modelos de K-medias y mezclas gaussianas. Algoritmo de maximización de expectativas. Selección del orden de los modelos: métodos BIC y silueta.

**Dedicación:** 2h 15m

Grupo grande/Teoría: 2h

Aprendizaje autónomo: 0h 15m



## SISTEMA DE CALIFICACIÓN

---

Cuestionarios: 50%

Competición: 50%

## NORMAS PARA LA REALIZACIÓN DE LAS PRUEBAS.

---

No pueden usarse calculadora, teléfonos móviles ni apuntes.

## BIBLIOGRAFÍA

---

### Básica:

- Aurelien Geron. Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. 2. UK: O'Reilly Media, 2022. ISBN 978-1098125974.

### Complementaria:

- Bishop, C.M. Pattern recognition and machine learning. New York: Springer, 2006. ISBN 0387310738.

- Kuncheva, L.I. Combining pattern classifiers: methods and algorithms [en línea]. 2nd ed. Hoboken (NJ): J. Wiley & Sons, 2014 [Consulta: 21/09/2018]. Disponible a: <https://onlinelibrary.wiley.com/doi/book/10.1002/9781118914564>. ISBN 9781118914564.

- Hastie, T.; Tibshirani, R.; Friedman, J.H. The elements of statistical learning: data mining, inference and prediction [en línea]. 2nd ed. New york [etc.]: Springer, 2009 [Consulta: 18/07/2023]. Disponible a: <https://link-springer-com.recursos.biblioteca.upc.edu/book/10.1007/978-0-387-84858-7>. ISBN 9780387848570.

## RECURSOS

---

### Otros recursos:

Cursos online

<https://es.coursera.org/learn/machine-learning?> />Competiciones

<http://www.kdd.org/kdd-cup> /> <http://www.kaggle.com/competitions> /> <https://datahack.analyticsvidhya.com/contest/all/> />