



# Guía docente

## 295601 - AB - Aprendizaje Bioestadístico

Última modificación: 30/01/2024

**Unidad responsable:** Escuela de Ingeniería de Barcelona Este  
**Unidad que imparte:** 749 - MAT - Departamento de Matemáticas.

**Titulación:** GRADO EN INGENIERÍA BIOMÉDICA (Plan 2009). (Asignatura optativa).

**Curso:** 2023      **Créditos ECTS:** 6.0      **Idiomas:** Castellano

### PROFESORADO

---

**Profesorado responsable:** Alférez Baquero, Edwin Santiago  
Mujica Delgado, Luis Eduardo  
Ruiz Ordoñez, Magda Liliana

**Otros:** Alférez Baquero, Edwin Santiago  
Ruiz Ordoñez, Magda Liliana  
Mujica Delgado, Luis Eduardo

### CAPACIDADES PREVIAS

---

Durante la etapa de formación de pregrado se han visto algunos fundamentos de la programación, así como ciertos conceptos matemáticos. En este curso se podrá apreciar algunas utilidades, específicamente:

1. Python: La utilización de variables, tipos de datos, declaraciones, expresiones, operadores y precedencia.
2. Conceptos Matemáticos: A lo largo del curso podrás ver la aplicación de algunos conceptos de estadística, cálculo y álgebra.
  - Estadística: Comprensión de medidas estadísticas básicas (media, mediana, desviación estándar), probabilidad condicional y prueba de hipótesis.
  - Cálculo: Conceptos fundamentales de diferenciación e integración, principalmente su relación con problemas de optimización en ciencia de datos.
  - Álgebra: Álgebra lineal básica, incluidas operaciones matriciales y espacios vectoriales, ya que son cruciales para comprender las estructuras de datos, la complejidad algorítmica y los modelos de aprendizaje automático.

Si bien una comprensión firme de estos conceptos matemáticos será útil para comprender el material del curso, el desempeño detallado de estas habilidades específicas no será parte de la evaluación del curso.

### REQUISITOS

---

Esta asignatura no presupone ningún requisito previo. Si es necesario tener un gran interés en comprender y aplicar métodos para el análisis de datos biomédicos. También son esenciales el entusiasmo y la voluntad por abordar los conceptos de las nuevas tecnologías para la solución de problemas desafiantes.

## METODOLOGÍAS DOCENTES

---

Este curso adopta un modelo de aprendizaje integrado, donde los conceptos teóricos y la aplicación práctica se entrelazan perfectamente durante cuatro horas semanales. Cada sesión se lleva a cabo dentro de un laboratorio de computación. Tras la introducción de cada concepto teórico, se explorará inmediatamente su implementación práctica a través de la programación en Python promoviendo un entorno de aprendizaje activo (teoría del aprendizaje constructivista). Al inicio de cada clase, se proporcionarán todos los materiales de Python. De esta manera, la atención se centra en aprender y aplicar nuevos conceptos.

Además, el curso aprovecha hasta cierto punto el modelo de aula invertida, asignando el 60% de la trayectoria de aprendizaje al estudio autónomo. Se espera que los estudiantes participen en un aprendizaje autodirigido, fomentando la investigación independiente y la consolidación de conocimientos. Este componente está diseñado para cultivar un entorno centrado en el alumno, permitiéndoles asumir la responsabilidad de su aprendizaje y desarrollar habilidades de autorregulación, que son fundamentales en el campo de la gestión de datos. La estrategia pedagógica distribuye las ponderaciones de la evaluación en consecuencia:

- Conocimientos basados en la teoría: 20%
- Tareas y proyectos de laboratorio de informática: 20%
- Aprendizaje autónomo: 60%

Como marco de evaluación, se desarrollarán tres trabajos destinados a fomentar el compromiso y la aplicación continua de los conceptos, garantizando una experiencia de aprendizaje holística que sea a la vez rigurosa y contextualmente relevante.

## OBJETIVOS DE APRENDIZAJE DE LA ASIGNATURA

---

Los objetivos están diseñados para fomentar una experiencia de aprendizaje holístico que combine conocimientos técnicos con habilidades sociales esenciales como el pensamiento crítico, la colaboración y la investigación independiente:

- Obtener una comprensión sólida de los principios fundamentales del aprendizaje automático en lo que respecta al campo de la ingeniería biomédica.
- Aprender a articular y enmarcar desafíos en diversos contextos biomédicos.
- Adquirir una comprensión integral de un espectro de los algoritmos para el procesamiento y análisis de datos y de imágenes, incluidos sus posibles beneficios y limitaciones.
- Aplicar estos nuevos conocimientos para abordar y resolver problemas biomédicos de complejidad moderada, fomentando el dominio de la codificación Python en contextos de análisis de datos y de imágenes.
- Desarrollar la capacidad de evaluación crítica de los resultados tomando decisiones metodológicas bien informadas.
- Fomentar el desarrollo de habilidades de aprendizaje autónomo, capacitando a los estudiantes para navegar y dominar diferentes campos de forma independiente.
- Cultivar un entorno de aprendizaje colaborativo donde se valore y promueva el intercambio de ideas y la participación constructiva del grupo.
- Inculcar una mentalidad impulsada por la investigación, impulsando al estudiantado a explorar más allá de los materiales proporcionados y sintetizar información de diversas fuentes.
- Dotar a los estudiantes de la capacidad de comparar críticamente diversas técnicas y recomendar las más adecuadas para determinadas cuestiones biomédicas.

## HORAS TOTALES DE DEDICACIÓN DEL ESTUDIANTADO

---

Tipo	Horas	Porcentaje
Horas grupo grande	30,0	20.00
Horas actividades dirigidas	90,0	60.00
Horas grupo pequeño	30,0	20.00

Dedicación total: 150 h

## CONTENIDOS

### 1. Introducción

#### Descripción:

Las sesiones están diseñadas para proporcionar una visión general del desarrollo histórico de los procesos industriales y establecer las bases para comprender el papel del procesamiento automatizado en el análisis contemporáneo de datos e imágenes dentro de las aplicaciones biomédicas. Se examinará la importancia de la adquisición de datos, el control de procesos y el monitoreo, junto con varios modelos utilizados para el análisis e interpretación de datos.

#### Objetivos específicos:

Familiarizarse con los fundamentos conceptuales del aprendizaje automático y su aplicación en la ingeniería biomédica. Esto incluye comprender la importancia de los datos para impulsar mejoras en los procesos, los conceptos básicos de la adquisición e instrumentación de datos y una introducción a los tipos de algoritmos que se encontrarán a lo largo del curso.

#### Actividades vinculadas:

Clases teóricas: Presentación del curso e Introducción.

Sesiones de laboratorio: Introducción a Python, Matplotlib, NumPy, Pandas y exploración de datos.

#### Dedicación: 10h

Grupo grande/Teoría: 2h

Grupo pequeño/Laboratorio: 2h

Aprendizaje autónomo: 6h

### 2. Aprendizaje supervisado

#### Descripción:

El aprendizaje supervisado es un tipo de aprendizaje automático en el que un algoritmo aprende a partir de datos de entrenamiento etiquetados y los utiliza para hacer predicciones o decisiones sin intervención humana. Se presentarán los fundamentos básicos de regresión y clasificación. Se estudiarán las principales técnicas empleadas para el modelo predictivo dentro del dominio de la ciencia de datos.

La regresión abarca la metodología estadística para modelar la relación entre una variable dependiente cuantitativa y una o más variables independientes. Se pondrá énfasis en la comprensión de la teoría, los supuestos subyacentes y la implementación práctica de modelos de regresión en el contexto del análisis de datos biomédicos.

En las sesiones de clasificación se enseñarán algoritmos destinados a la categorización de datos en clases predeterminadas. Se examinarán una variedad de técnicas de clasificación. Los ejercicios prácticos facilitarán la comprensión de la aplicación de estos métodos en la clasificación de conjuntos de datos biomédicos.

#### Objetivos específicos:

Dotar las habilidades necesarias para construir, interpretar y evaluar modelos predictivos utilizando regresión y algoritmos de clasificación. Se centrará en la aplicación de estos modelos a conjuntos de datos biomédicos, la interpretación de los resultados con ojo crítico y comprensión de las implicaciones de los modelos en escenarios biomédicos del mundo real. Este conocimiento servirá como trampolín para temas más avanzados y la toma de decisiones.

#### Actividades vinculadas:

Clases teóricas y sesiones de laboratorio: Regresión Lineal, Bayes-Naives, Regresión Logística, Análisis discriminante, K vecinos cercanos (KNN), Máquinas de Soporte Vectorial (SVM). Índices para la evaluación de modelos y matriz de confusión.

Trabajo 1.

#### Dedicación: 40h

Grupo grande/Teoría: 8h

Grupo pequeño/Laboratorio: 8h

Aprendizaje autónomo: 24h

### 3. Remuestreo y aprendizaje en conjunto

**Descripción:**

Esta parte del curso presentará metodologías avanzadas esenciales para mejorar el rendimiento del modelo predictivo. Se exploran estrategias de remuestreo (resampling), como la validación cruzada (cross-validation), bootstrapping y los conjuntos de validación, por su papel fundamental en la evaluación, selección y estimación de la precisión de los modelos. Los métodos de aprendizaje conjunto (Ensemble Learning) abarcan enfoques integradores como embolsado (bagging), impulso (boosting) y apilamiento (stacking), que combinan múltiples modelos para superar la precisión y estabilidad de modelos predictivos singulares.

**Objetivos específicos:**

Impartir una comprensión integral de las técnicas de remuestreo para una validación sólida y métodos de aprendizaje en conjunto para aumentar la efectividad del modelo. La aplicación de estas metodologías se demostrará utilizando Python en conjuntos de datos biomédicos, proveyendo con las habilidades necesarias para mitigar el sobreajuste, la variación y el sesgo en el desarrollo de modelos. Dominar estas técnicas es crucial para desarrollar modelos avanzados y generalizables en el análisis de datos.

**Actividades vinculadas:**

Clases teóricas y sesiones de laboratorio: Cross-validation, Bootstrapping, Ensemble Learning, bagging, boosting, stacking, árboles de decisión.

**Dedicación:** 20h

Grupo grande/Teoría: 4h

Grupo pequeño/Laboratorio: 4h

Aprendizaje autónomo: 12h

### 4. Aprendizaje no supervisado

**Descripción:**

En estas sesiones se profundizará en la exploración de estructuras de datos sin resultados preetiquetados. Se prestará especial atención a las técnicas de agrupación, como las k-medias y la agrupación jerárquica, que agrupan puntos de datos en función de métricas de similitud. Se examinará la utilidad de los métodos de reducción de dimensionalidad, como lo es el análisis de componentes principales (PCA) para simplificar conjuntos de datos complejos y así revelar patrones intrínsecos.

**Objetivos específicos:**

Aplicar técnicas de aprendizaje no supervisadas de manera efectiva para discernir estructuras subyacentes en datos biomédicos. Esto incluye identificar agrupaciones naturales dentro de conjuntos de datos, reducir la complejidad de los datos de alta dimensión para una mejor visualización e interpretación y preparar los datos para análisis posteriores. A través de esto, se adquirirá competencia para revelar patrones y tendencias.

**Actividades vinculadas:**

Clases teóricas y sesiones de laboratorio: k-means, Density-Based Spatial Clustering of Applications with noise (DBSCAN), hierarchical clusterin, Principal Component Analysis (PCA).

Trabajo 2.

**Dedicación:** 20h

Grupo grande/Teoría: 4h

Grupo pequeño/Laboratorio: 4h

Aprendizaje autónomo: 12h



## 5. Introducción al aprendizaje profundo

### Descripción:

Esta sección del curso abordará el aprendizaje profundo, poniendo especial énfasis en las redes neuronales y sus diversas arquitecturas. Se introducirán conceptos fundamentales como las redes neuronales convolucionales (CNN) y las redes neuronales recurrentes (RNN), además de técnicas esenciales como el ajuste de parámetros y la regularización. Se explorará el uso práctico de las redes profundas en la ingeniería biomédica, particularmente en la clasificación de imágenes y de documentos, y en la predicción de series temporales. El estudiantado aprenderá a implementar y entrenar modelos de aprendizaje profundo usando Python y bibliotecas populares de aprendizaje profundo.

### Objetivos específicos:

Impartir una comprensión sólida de la estructura y funcionamiento de las redes neuronales artificiales, permitiendo la implementación práctica de estos modelos para resolver problemas de ingeniería biomédica del mundo real. Los participantes adquirirán las habilidades para diseñar y entrenar arquitecturas de redes neuronales, comprenderán sus ventajas y limitaciones y aprovecharán este conocimiento para desarrollar soluciones efectivas para desafíos complejos de análisis de datos biomédicos. Esta fundación tiene como objetivo allanar el camino para estudios avanzados en aprendizaje automático y aplicaciones innovadoras dentro del campo biomédico.

### Actividades vinculadas:

Clases teóricas y sesiones de laboratorio: Estructura de redes neuronales (neuronas, capas y funciones de activación). Algoritmos con redes Feedforward and backpropagation. Diferentes tipos de arquitecturas de redes neuronales como Perceptrones Multicapa (MLP), redes neuronales convolucionales (CNN) y redes neuronales recurrentes (RNN). Estrategias para prevenir el sobreajuste, como métodos de regularización y abandono (dropout).

### Dedicación: 40h

Grupo grande/Teoría: 8h

Grupo pequeño/Laboratorio: 24h

Aprendizaje autónomo: 8h

## 6. Introducción a los Transformers

### Descripción:

Este módulo final del curso introduce la arquitectura de red neuronal profunda Transformers, iniciando por el módulo de atención y estudiando los fundamentos de la arquitectura Transformer original de codificador y decodificador (encoder-decoder). Se explorarán las arquitecturas básicas derivadas del codificador como BERT, y de la parte decodificador como GPT. También, introducirá una modificación del Transformer para aplicarlo en imágenes biomédicas, denominada Transformadores de Visión (ViT). Se explorarán estas arquitecturas avanzadas y se comprenderán los principios fundamentales que impulsan su desempeño y las formas innovadoras en que se aplican en la resolución de problemas complejos. El módulo también proporcionará una exposición práctica para implementar estos algoritmos utilizando marcos modernos de aprendizaje profundo y demostrar su aplicación.

### Objetivos específicos:

Proveer una comprensión profunda de las arquitecturas de aprendizaje profundo avanzadas y sus aplicaciones dotando de las habilidades necesarias no solo para comprender los fundamentos teóricos de estos modelos sofisticados. El dominio de estas técnicas avanzadas es crucial para contribuir significativamente al campo del análisis de datos biomédicos, impulsando la innovación y ampliando los límites de lo que es posible con la IA en la atención médica y la medicina.

### Actividades vinculadas:

Clases teóricas y prácticas de laboratorio para entender y aplicar la arquitectura de los Transformers, comenzando por el estudio del mecanismo de atención y cómo se utiliza en la estructura encoder-decoder. Se analizarán y pondrán en práctica arquitecturas derivadas del Transformer como BERT para el procesamiento del lenguaje natural y GPT para generación de texto, así como la adaptación de Transformers al campo de la visión por computadora con los ViT. Aprenderemos sobre la importancia de la atención y cómo ha revolucionado el análisis de datos secuenciales. Los estudiantes tendrán la oportunidad de entrenar modelos de Transformer en tareas de clasificación de imágenes y comprensión del lenguaje natural, utilizando frameworks de aprendizaje profundo actuales. Se alentará a los estudiantes a experimentar con estos modelos avanzados para desarrollar soluciones innovadoras a problemas de análisis de datos biomédicos.

Trabajo 3

### Dedicación: 20h

Grupo grande/Teoría: 4h

Grupo pequeño/Laboratorio: 4h

Aprendizaje autónomo: 12h

## SISTEMA DE CALIFICACIÓN

La evaluación se realizará de forma continua durante el desarrollo de cada módulo mediante trabajos prácticos.

## NORMAS PARA LA REALIZACIÓN DE LAS PRUEBAS.

## BIBLIOGRAFÍA

### Básica:

- Bishop, Christopher M. Pattern recognition and machine learning. Springer, 2006. ISBN 9780387310732.
- Géron, Aurélien. Hands-on machine learning with scikit-learn & tensorflow : concepts, tools, and techniques to build intelligent systems [en línea]. Sebastopol, CA: O'Reilly Media, Inc, 2017 [Consulta: 15/04/2020]. Disponible a: <https://ebookcentral.proquest.com/lib/upcatalunya-ebooks/detail.action?docID=4822582>. ISBN 9781491962268.
- James, G.; Witten, D.; Hastie, T.; Tibshirani, R. An introduction to statistical learning with applications in R. Springer, 2013. ISBN 9781461471370.
- Raschka, Sebastian. Python machine learning : machine learning and deep learning with Python, scikit-learn, and TensorFlow [en línea]. 2nd ed. Birmingham, UK: Packt Publishing Ltd, 2017 [Consulta: 14/04/2020]. Disponible a: <https://ebookcentral.proquest.com/lib/upcatalunya-ebooks/detail.action?docID=5050960>. ISBN 9781787126022].



**Complementaria:**

- Hastie, Trevor; Tibshirani, Robert; Friedman, Jerome. The Elements of statistical learning : data mining, inference, and prediction [en línea]. 2nd ed. New York, NY: Springer Series in Statistics, 2001 [Consulta: 27/08/2018]. Disponible a: <http://dx.doi.org/10.1007/978-0-387-84858-7>. ISBN 9780387848587.

## RECURSOS

---

**Otros recursos:**

Material disponible en ATENEA por parte de los responsables del curso.