



PRESENTACIÓN

Machine Learning II

- **Breve descripción de la asignatura:** Este curso se centra en las técnicas de minería de datos, con aplicación a la biología de sistemas. En particular, se cubrirá el área de la teoría del aprendizaje, con énfasis en el bias/variance trade-off que regulan el overfitting y el underfitting, y en la regresión regularizada. Veremos también los algoritmos de aprendizaje no supervisado K-means y Gaussian Mixture Models (GMM), así como el algoritmo de Expectation-Maximization (EM) con aplicaciones variadas. Cubriremos también técnicas para seleccionar los modelos más adecuados en función del data disponible, así como técnicas de reducción de dimensionalidad que pueden ayudar en casos en lo que no se tenga acceso a muchos datos, así como facilitar la visualización de los datos. Finalmente, se hará una introducción al aprendizaje con Deep Learning. El curso concluirá con un caso práctico de minería de datos aplicada al Single-cell RNA-Sequencing (scRNA-seq).
- **Carácter:** Optativa
- **ECTS:** 3
- **Curso y semestre:** Primer curso y segundo semestre
- **Idioma:** Español. Se requieren conocimientos de inglés
- **Título:** Máster en Métodos Computacionales en Ciencias
- **Módulo y materia de la asignatura:** Módulo 4 Optativo y Materia 4.1 Optatividad
- **Profesor responsable de la asignatura:** Mikel Hernández Arrazola
- **Profesores:** Mikel Hernández Arrazola e Idoia Ochoa Álvarez
- **Horario:** [Calendario del Máster](#)
- **Aula:** 1 edificio Los Castaños

COMPETENCIAS

| | |
|---------------------|--|
| MACHINE LEARNING II | |
| | COMPETENCIAS BÁSICAS |
| CB6 | Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo o aplicación de ideas, a menudo en un contexto de investigación. |



| | |
|------|--|
| CB7 | Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio. |
| CB10 | Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo. |
| | COMPETENCIAS GENERALES |
| CG3 | Conocer los principales problemas que se presentan en la adquisición y tratamiento de datos experimentales y cómo darles respuesta. |
| CG4 | Comunicar tanto de manera oral como escrita un tema o datos de investigación en el área de las ciencias experimentales. |
| | COMPETENCIAS ESPECÍFICAS |
| CE5 | Aplicar los métodos computacionales de procesamiento de datos a un problema científico particular de la disciplina de interés para el estudiante. |
| CE6 | Diseñar un experimento científico para que sea rico en información, recogiendo gran cantidad de datos de manera estructurada que faciliten su procesamiento posterior. |
| CE8 | Adquirir datos (bien en el laboratorio, o bien mediante minería on-line), organizarlos, filtrarlos, procesarlos, representarlos y refinarlos. |
| CE9 | Extraer información de los datos con técnicas computacionales siguiendo un método científico. |



| | |
|-------|--|
| CE10 | Presentar los datos experimentales y la información científica de manera que se comuniquen de manera eficiente y fidedigna. |
| | COMPETENCIAS ESPECÍFICAS DE OPTATIVIDAD |
| CEOP8 | Conocer los principales métodos de acceso a fuentes externas de datos (principalmente biológicos) y su visualización y análisis mediante teoría de grafos. |

PROGRAMA

Teoría del aprendizaje:

- Bias/Variance trade-off
- Regresión regularizada

Mixtura de Models Gaussianos:

- K-means
- Metodo de Expectation Maximization

Caso practico: Minería de datos para el descubrimiento de biomarcadores asociados al cancer de mama

Selección de modelos:

- Cross-validation
- Selección de variables

Reducción de dimensionalidad:

- Principal Component Analysis (PCA)
- tSNE y UMAP

Introducción al aprendizaje profundo: Deep Learning

Caso práctico: Minería de datos aplicada al Single-cell RNA-Sequencing

ACTIVIDADES FORMATIVAS

ACTIVIDADES FORMATIVAS:

Asignatura de 3 ECTS que equivalen a 75 horas.

1.- ACTIVIDADES PRESENCIALES (30 horas)

- Clases presenciales teóricas (20 horas)



- Clases presenciales prácticas (10 horas)

2.- ACTIVIDADES NO PRESENCIALES (45 horas)

- Trabajos dirigidos (15 horas)

- Estudio personal (30 horas)

EVALUACIÓN

La evaluación de este modulo será continua. Los alumnos se expondrán a las siguientes evaluaciones:

- Casos prácticos (30%). Entrega en la ultima clase.
- Trabajo final de modulo y exposición publica del trabajo (50% y 20% respectivamente). Esta actividad se realizara en la ultima clase.

HORARIOS DE ATENCIÓN

No hay horario específico para la atencion de estudiantes. Concertar cita previa por e-mail y se acuerda la fecha y hora de la tutoría:

- Profesor responsable de la asignatura: Mikel Hernaez Arrazola
- Profesores:
 - Idoia Ochoa Alvarez, email: iochoal@unav.es
 - Mikel Hernández Arrazola, email: mhernaez@unav.es

BIBLIOGRAFÍA

Bibliografía básica:

- An Introduction to Statistical Learning. Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani, 2018 ISBN-13: 978-1461471370 [Localízalo en la Biblioteca](#) [recurso electrónico]
- **Machine Learning with R: Expert Techniques for Predictive Modeling to Solve All Your Data Analysis Problems.** Lantz, Brett. Birmingham: Packt Publishing. 2015. [Localízalo en la Biblioteca](#)
- **Cluster and Classification Techniques for the Biosciences.** Fielding, Alan. Cambridge, UK : Cambridge University Press. 2007. [Localízalo en la Biblioteca](#) [recurso electrónico]

Bibliografía recomendada:

- **Genome Data Analysis.** Kim, Ju Han. Singapore : Springer 2019. XVI, 367 p. : 645 il., 236 il. col. [Localízalo en la Biblioteca](#) [recurso electrónico]
- **Introduction to machine learning with Python: a guide for data scientists.** Andreas C. Müller and Sarah Guido. Müller, Andreas C.. Sebastopol, CA : O'Reilly, cop. 2017. XII, 378 p. : il. [Localízalo en la Biblioteca](#)
- **Python machine learning** Lee, Wei-Meng. Indianapolis, IN : Wiley, [2019] 1 recurso electrónico (xxiv, 296 p.) [Localízalo en la Biblioteca](#) [recurso electrónico]

Artículos:

- *Opportunities and obstacles for deep learning in biology and medicine.* Ching, Travers, Daniel S. Himmelstein, Brett K. Beaulieu-Jones, Alexandr A. Kalinin, Brian



Universidad
de Navarra

- T. Do, Gregory P. Way, Enrico Ferrero et al. *Journal of The Royal Society Interface* 15, no. 141 (2018): 20170387. [10.1098/rsif.2017.0387](https://doi.org/10.1098/rsif.2017.0387) [Localízalo en la Biblioteca](#)
- *Machine learning for integrating data in biology and medicine: Principles, practice, and opportunities.* Marinka Zitnik, Francis Nguyen, Bo Wang, Jure Leskovec, Anna Goldenberg, Michael M. Hoffman <https://doi.org/10.1016/j.inffus.2018.09> [Localízalo en la Biblioteca](#)
 - *Next-generation machine learning for biological networks.* Camacho, Diogo M., Katherine M. Collins, Rani K. Powers, James C. Costello, and James J. Collins. *Cell* 173, no. 7 (2018): 1581-1592. <https://doi.org/10.1016/j.cell.2018.05.015> [Localízalo en la Biblioteca](#)