



## PRESENTACIÓN

### Breve descripción:

Es evidente que la biología y las ciencias médicas son dos disciplinas emergentes en el uso de la ciencia de datos. Aunque los algoritmos de *machine learning* funcionan bastante bien para las predicciones en numerosos campos como la economía, matemáticas o las ciencias sociales, por el contrario, en la biología aún tienen mucho que mejorar dada la enorme complejidad de los sistemas vivos. En la actualidad estamos viendo el surgimiento de las ciencias *ómicas* que engloban la captura e interpretación de datos biológicos en multitud de dimensiones creando la necesidad de integrar datos como la estructura y composición del genoma, marcadores genéticos, la regulación de la expresión génica, la estructura de proteínas y su efecto sobre el genoma y transcriptoma para poder entender el funcionamiento de organismos complejos como el ser humano.

En este módulo sentaremos las bases sobre los conceptos de machine learning usados en las ciencias de datos biológicas. Para ello introduciremos conceptos de modelos estadísticos tradicionales como contrastes de hipótesis, modelos lineales, cálculo de significación por ANOVA, regresión y modelos generalizados. Posteriormente introduciremos conceptos de aprendizaje automático, clasificaciones supervisadas y no supervisadas así como modelos de tomas de decisión. Finalmente usaremos prácticamente estos conceptos para aplicarlos al análisis de expresión diferencial de genes así como los últimos avances en cuantificación y variación a nivel de transcrito.

- **Titulación:** Máster en Métodos Computacionales en Ciencias
- **Módulo/Materia:** Módulo 4: optativo / Materia 4.1: optatividad
- **ECTS:** 3
- **Curso, semestre:** 1º curso 2º semestre
- **Carácter:** Optativa
- **Profesorado:**
  - Jesús Fernando López Fidalgo, email: fidalgo(at)unav.es
  - Mikel Hernáez Arrazola, email: mhernaez(at)unav.es
  - Ibon Tamayo Uria, email: itamayou(at)unav.es
- **Idioma:** Español. Se requieren conocimientos de inglés.
- **Aula, Horario:** 1 edificio Los Castaños. [Calendario del Máster](#)

## COMPETENCIAS

MACHINE LEARNING I	
	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.
	<b>COMPETENCIAS GENERALES</b>
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.
	<b>COMPETENCIAS ESPECÍFICAS</b>
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.
	<b>COMPETENCIAS ESPECÍFICAS DE OPTATIVIDAD</b>
CEOP6	Implementar clasificadores en diversos ámbitos científicos, entendiendo los fundamentos de su entrenamiento y de su evaluación mediante las métricas más adecuadas.

## PROGRAMA

Contrastes de hipótesis:

- Concepto
- Contrastes de medias y proporciones
- Tablas de contingencia

Introducción a los modelos lineales:

- ANOVA: diseño experimental, interacciones, contrastes post hoc múltiples, análisis de residuos
- Regresión
- Modelos lineales generalizados

Introducción a los modelos de aprendizaje automático supervisado: Regresión

- Análisis frecuentista vs. Bayesiano
- Regresión regularizada
- Regression lineal desde el punto de vista Bayesiano

Introducción a los modelos de aprendizaje automático supervisado: Clasificación

- Regresión logística
- Support Vector Machine



# Universidad de Navarra

- Árboles de decisión y métodos de conjunto

Introducción a los modelos de aprendizaje automático no supervisado

- Reducción de dimensionalidad: PCA
- Clusterización: K-means

El dogma central de la biología molecular:

- Tecnologías transcriptómicas de secuenciación
- Cuantificación de la expresión génica
- Clusterización y caracterización de muestras, hierarchical y k-means
- Reducción de dimensiones PCA
- Análisis de expresión diferencial: teoría y práctica
- Integración y anotación de datos
- Análisis funcional: Gene Set Enrichment Analysis
- Corrección de desviación por lotes
- Cuantificación a nivel de transcrito, nuevos algoritmos sin mapeo genómico

## 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 módulo será continua. Los alumnos se expondrán a las siguientes evaluaciones:

- Prueba escrita teoría y práctica con R (40%, Jesús López Fidalgo).
- Puesta en práctica en R (30% Ibon Tamayo Uria).
- Trabajo final de módulo y exposición pública del trabajo (30%, Ibon Tamayo Uria).

## HORARIOS DE ATENCIÓN



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

- Profesores:
  - Jesús Fernando López Fidalgo, email: fidalgo(at)unav.es
  - Mikel Hernáez Arrazola, email: mhernaiez(at)unav.es
  - Ibon Tamayo Uria, email: itamayou(at)unav.es

## BIBLIOGRAFÍA

### Bibliografía básica:

- [An Introduction to Statistical Learning](#). Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani, 2018. ISBN : 1-4614-7138-9 [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]
- **Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2**. Love, M.I., Huber, W., Anders, S. (2014) Genome Biology, 15:550. [10.1186/s13059-014-0550-8](#) [Localízalo en la Biblioteca](#)
- [DeSeq2 Tutorial](#)

### 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. 2016. 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]

*Detalle disponible únicamente*

### Artículos:

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