



MACHINE LEARNING I

PRESENTACIÓN

Breve descripción

Este curso sentaremos las bases sobre los conceptos de *machine learning* usados en la ciencia de datos. A través de una combinación de teoría y ejercicios prácticos, los estudiantes explorarán algoritmos clave, técnicas de modelado y métodos de aprendizaje automático aplicados en diversos campos, desde la biomedicina hasta el análisis de datos complejos. Los participantes no solo desarrollarán competencias teórico-técnicas en herramientas como Python y Keras, sino que también aprenderán a abordar problemas reales trabajando en casos de uso con datos reales. 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 a la estratificación automática de pacientes de cancer de mama (no se requiere ningún conocimiento previo de biomedicina).

- **Titulación:** MÁSTER UNIVERSITARIO EN CIENCIA DE DATOS PARA CIENCIAS EXPERIMENTALES
- **Módulo/Materia:** Módulo 2/Materia 2.3. Fundamentos de Aprendizaje Automático
- **ECTS:** 5
- **Curso, semestre:** 1º curso. 1º semestre
- **Carácter:** Obligatoria
- **Profesorado:**
Iñaki Echeverría Huarte, email: iecheverriah@unav.es
Ibon Tamayo Uria, email: itamayou@unav.es
- **Idioma:** Español. Se requieren conocimientos de inglés.
- **Aula, Horario:** Consultar el calendario del máster.

RESULTADOS DEL APRENDIZAJE

RA1 Aplicar soluciones computacionales para problemas científicos, usando una amplia gama de plataformas de análisis de datos

RA6 Evaluar las implicaciones éticas de los sistemas automáticos e inteligencia artificial en las ciencias experimentales, considerando sesgos algorítmicos, discriminación y transparencia en la toma de decisiones

RA7 Implementar algoritmos eficientes para la resolución de problemas científicos utilizando programación estructurada y pseudocódigo

RA12 Manejar con soltura el lenguaje de programación R para la manipulación y análisis de datos científicos.

RA13 Implementar las nociones teóricas de la ciencia de datos en un entorno computacional utilizando el lenguaje de programación R.

RA14 Manipular y procesar datos científicos de diversas fuentes (archivos de texto, bases de datos, etc.) utilizando técnicas de data wrangling en diferentes entornos de programación

RA15 Generar e interpretar representaciones gráficas adecuadas para comunicar resultados científicos de manera clara y efectiva en diferentes entornos de programación



RA17 Implementar técnicas de inferencia estadística para analizar datos científicos, extraer conclusiones y tomar decisiones informadas.

RA 18 Desarrollar modelos de aprendizaje automático de extremo a extremo para resolver problemas de predicción, clasificación y agrupamiento en ciencias experimentales.

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
- Árboles de decisión y métodos de conjunto

Introducción a los modelos de aprendizaje automático no supervisado

- Reducción de dimensionalidad: PCA, tSNE y UMAP
- Clusterización: K-means
- Mixtura de Models Gaussianos
- Metodo de Expectation Maximization
- Caso practico: Minería de datos para el descubrimiento de biomarcadores asociados al cancer de mama

End-to-end machine learning:

- Selecccion de modelos
- Bias/Variance trade-off - Overfitting
- Cross-validation
- Selecccion de variables
- Caso practico

ACTIVIDADES FORMATIVAS

Asignatura de 5 ECTS que equivalen a **50 horas presenciales**.

1.-ACTIVIDADES PRESENCIALES (50 horas)

- Clases presenciales teóricas (30 horas)
- Clases presenciales prácticas (20 horas)

2.-ACTIVIDADES NO PRESENCIALES (75horas)

- Trabajos dirigidos (15 horas)
- Estudio personal (60 horas)



Universidad
de Navarra

Facultad de Ciencias

EVALUACIÓN

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

- Intervención en clases, seminarios y clases prácticas (10%)
- Prueba escrita teoría y practica con R - Aprendizaje automatico supervisado (30%)
- Trabajo en práctica en R – Aprendizaje automatico no supervisado (20%)
- Trabajo en práctica en R – End-to-end machine learning (10%)
- Trabajo final de modulo y exposición pública del trabajo (30%)

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 López Fidalgo, email: fidalgo@unav.es
- Idoia Ochoa, email: iochoal@unav.es
- Ibon Tamayo, email: itamayou@unav.es

BIBLIOGRAFÍA

- [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](#)
- 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]