

Breve descripción de la actividad

Introducción al *Machine Learning* en biomedicina. Las clases están enfocadas en conocer los pasos necesarios para procesar diferentes fuentes de datos biomédicos, buscar patrones, extraer conclusiones e implementar modelos predictivos.

La asignatura tiene un gran componente práctico en el que los asistentes aprenderán nociones básicas de programación en *Python* y *scikit-learn* para poder implementar las diferentes herramientas necesarias para el procesamiento de los datos.

No se requieren conocimientos previos de programación.

Objetivos formativos

- Conocer problemas prácticos en biomedicina que han podido resolverse gracias a distintas técnicas de ML.
- Ser capaz de diseñar e implementar un modelo para la predicción/clasificación de datos biomédicos.
- Aprender a extraer patrones que agrupen diferentes categorías en los datos.
- Conocer los algoritmos del estado del arte para predecir/clasificar datos.
- Aprender a detectar problemas y posibles soluciones para mejorar la predicción del modelo diseñado.
- Aprender a usar la herramienta *scikit-learn* para analizar los datos.

Profesorado

Javier Rasero Daparte
IIS Biocruces
Teléfono 94.600.6000 ext 5298
Email: jrasero.daparte@gmail.com

Inscripción

Es necesario enviar un correo electrónico a la dirección: aitziber.mojas@ehu.eus (Aitziber Mojas), adjuntando el Anexo I completado e indicando en el asunto: *Curso Investigación Biomédica*. El archivo adjunto debe guardarse con el Nombre y Apellidos de la persona que se inscribe.

La asistencia es obligatoria para conseguir el *Certificado* del curso, que deberá incluirse en el documento de Actividades Formativas 2019 y subirse a la aplicación GAUR.

El plazo para la inscripción en la actividad finaliza el día **7 de marzo de 2019**. Posteriormente, se comunicará la admisión o no en el curso.

Dado que hay un número limitado de plazas, para la aceptación en el curso se tendrá en cuenta la fecha de inscripción y la pertenencia al programa de doctorando *Investigación Biomédica*.

Personas de contacto

Preinscripción
Aitziber Mojas
Teléfono: 946015683
Email: aitziber.mojas@ehu.eus

Responsable de la actividad
M^a Begoña Ruiz Larrea
Teléfono: 946012829
Email: mbego.ruizlarrea@ehu.eus

Actividades Formativas Específicas de Doctorado
Curso ofertado por el Programa de Doctorado
Investigación Biomédica

**Análisis, predicción
y clasificación de
datos biomédicos**

Inicio el 13 de marzo de 2019

TEMARIO

Tema I: Introducción al *Machine Learning* (ML)

- ¿Qué es el ML?
- Tipos de algoritmos: aprendizaje supervisado vs no supervisado.
- Ejemplos de uso de ML en el ámbito biomédico.
- Clase práctica: Introducción al lenguaje de programación *Python* y al entorno de *scikit-learn*.
- Clase práctica: Manipulación y visualización de los datos usando las librerías *pandas* y *matplotlib*.

Tema II: Aprendizaje Supervisado I. Regresión

- Predicción de datos continuos (regresión).
- Regresión lineal.
- Gradiente descendiente.
- Regresión polinomial.
- Métodos de regularización.
- Evaluación de los modelos de regresión.
- Clase práctica: Preparación de los datos. Construcción de un modelo de regresión para predicción de valores continuos.

Tema 3: Aprendizaje Supervisado II. Clasificación

- Clasificación de datos (identificación de la categoría de los datos).
- Regresión logística.
- *Support Vector Machines*.
- Árboles de decisión y *Random Forest*.
- *Lazy Learning algorithms*.
- Clasificación multiclase.
- Métricas de evaluación de modelos de clasificación.
- Clase práctica: Uso de *scikit-learn* para el aprendizaje, clasificación y predicción de datos. Comparación de modelos predictivos.
- Clase práctica: pre-procesamiento de los datos usando *scikit-learn*.

Tema 4: Selección del modelo e hiperparámetros y su evaluación

- *Cross-validation*.
- Curvas de aprendizaje y validación.
- *Grid Search*.
- Soluciones para reducir los errores del modelo.
- Clase práctica: Uso de *scikit-learn* para evaluar diferentes modelos y estimar los parámetros óptimos.
- Clase práctica: Mejorar y automatizar los modelos en *scikit-learn*: *Feature selection* y módulo *Pipeline*.

Tema 5: Aprendizaje no Supervisado

- Clustering (*k-means*, *Fuzzy c means*).
- Clustering jerárquico.
- DBSCAN.
- Evaluación de los clusters.
- Selección del número óptimo de clusters.
- Reducción de dimensionalidad.
- Clase práctica: Uso de algoritmos de clustering y de reducción de dimensionalidad en datos biomédicos.

Metodología

Las clases consisten en 2 partes fundamentales: Teoría y Prácticas (ver Calendario).

Local

Las clases tendrán lugar en el aula de informática P6 (primera planta) de la Facultad de Medicina y Enfermería de la UPV/EHU, Campus de Bizkaia.

El programa informático que se utiliza es Anaconda <https://www.continuum.io/downloads>, que no requiere licencia (se puede instalar en cualquier sistema operativo).

Calendario

La asignatura se imparte durante 5 semanas, un tema por semana, 5 horas semanales (25 horas).

Inicio: 13 de marzo. Fin: 11 de abril

Las clases están organizadas de la siguiente manera:

- Miércoles (2 h): 2 h teóricas para conocer cómo funcionan las diferentes herramientas a utilizar.
- Jueves (3 h): Se explicará cómo implementar y usar las diferentes herramientas necesarias haciendo uso de *scikit-learn* y se aplicarán a la resolución de un problema.

Lunes	Martes	Miércoles	Jueves	Viernes
Marzo 2019				
Semana 1: Introducción al <i>Machine Learning</i> y preparación de los datos				
11	12	13 16:00-18:00 Teoría	14 16:00-19:00 Prácticas	15
Semana 2: Aprendizaje Supervisado I: Regresión				
18	19	20 16:00-18:00 Teoría	21 16:00-19:00 Prácticas	22
Semana 3: Aprendizaje Supervisado II: Clasificación				
25	26	27 16:00-18:00 Teoría	28 16:00-19:00 Prácticas	29
Abril 2019				
Semana 4: Selección del modelo e hiperparámetros y su evaluación				
1	2	3 16:00-18:00 Teoría	4 16:00-19:00 Prácticas	5
Semana 5: Aprendizaje no Supervisado (<i>Clustering</i> y Reducción de dimensionalidad)				
8	9	10 16:00-18:00 Teoría	11 16:00-19:00 Prácticas	12

Anexo I. Hoja de inscripción en el curso (hasta el 7 de marzo, inclusive):

Análisis, predicción y clasificación de datos biomédicos

Datos Personales:

Nombre y Apellidos: _____

DNI: _____

Email de contacto: _____

Teléfono de contacto: _____

Programa de Doctorado en el que estás matriculado/a:

☐ Investigación Biomédica

☐ Otros: especificar _____

Año/Curso en el que realizaste la primera matrícula en el programa de doctorado _____

Director/es de la tesis (Nombre Apellidos): _____

Estoy interesado/a en participar en la siguiente actividad

Debido a que el número de plazas es limitado, la asistencia será obligatoria para conseguir el certificado de asistencia al curso. Éste deberá incluirse en el documento de Actividades Formativas y subirse a la aplicación GAUR.

Se tendrá en cuenta la fecha de inscripción al curso y la pertenencia al programa **Investigación Biomédica**.

Guarda la Hoja de inscripción con tu Nombre y Apellidos y envíala a la dirección:
aitziber.mojas@ehu.eus, indicando en el asunto **Curso Investigación Biomédica**.

El plazo para la inscripción en la actividad finaliza el día **7 de marzo de 2019**.

Posteriormente, se comunicará la admisión o no al curso.