

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 *sci-kit* para analizar los datos.

Profesorado

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

Jesús M. Cortés
IIS Biocruces
Teléfono 94.600.6000 ext 5199
Email: jesus.m.cortes@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 2018 y subirse a la aplicación GAUR.

El plazo para la inscripción en la actividad finaliza el día **18 de mayo de 2018**. 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 5 de Junio de 2018

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.
- Pre-procesado de los datos.
- Clase práctica: Introducción a las herramientas a utilizar para el análisis de datos, al lenguaje de programación Python y al entorno *sci-kit*.

Tema II: Aprendizaje Supervisado I. Regresión

- Predicción de datos continuos (regresión).
- Regresión lineal.
- Gradiente descendiente.
- Regresión polinomial.
- Evaluación de los modelos de regresión.
- Métodos de regularización.
- 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.
- Métricas de evaluación de modelos de clasificación.
- Clasificación multiclase.
- Clasificación multilabel.
- *Support Vector Machines*.
- Árboles de decisión.
- *Random Forest*.
- *Lazy Learning algorithms*.
- Clase práctica: Uso de *sci-kit* para el aprendizaje, clasificación y predicción de datos. Comparación de modelos predictivos.

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.
- *Ensemble Learning*.
- Clase práctica: Uso de *sci-kit* para evaluar diferentes modelos y estimar los parámetros óptimos.
- Clase práctica: Pasos para mejorar el modelo y la predicción.

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 P3 (primera planta) de la Facultad de Medicina y Enfermería de la UPV/EHU, Campus de Bizkaia.

Los ordenadores deben tener instalado el programa 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: 5 de Junio. Fin: 3 Julio

Las clases están organizadas de la siguiente manera:

- Lunes (2 h): 2 h teóricas para conocer cómo funcionan las diferentes herramientas a utilizar.
- Martes (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 |
|--|-----------------------------|--------------------------------|--------|---------|
| Junio 2018 | | | | |
| Semana 1: Introducción al <i>Machine Learning</i> y preparación de los datos | | | | |
| 4 | 5 16:00-18:00 Teoría | 6 16:00-19:00 Prácticas | 7 | 8 |
| Semana 2: Aprendizaje Supervisado I: Regresión | | | | |
| 11 | 12 16:00-18:00 Teoría | 13 16:00-19:00 Prácticas | 14 | 15 |
| Semana 3: Aprendizaje Supervisado II: Clasificación | | | | |
| 18 | 19 16:00-18:00 Teoría | 20 16:00-19:00 Prácticas | 21 | 22 |
| Semana 4: Selección del modelo e hiperparámetros y su evaluación | | | | |
| 25 | 26 16:00-18:00 Teoría | 27 16:00-19:00 Prácticas | 28 | 29 |
| Julio 2018 | | | | |
| Semana 5: Aprendizaje no Supervisado (<i>Clustering</i> y Reducción de dimensionalidad) | | | | |
| 2 | 3 16:00-18:00 Teoría | 4 16:00-19:00 Prácticas | 5 | 6 |