

Formulario de Curso

Asignatura:

Implementación de modelos basados en aprendizaje automático para el abordaje de problemas biológicos

Instituto o Unidad:

Instituto Pasteur Montevideo / Instituto de Investigaciones Biológicas Clemente Estable.

Departamento o Área:

Unidad de Bioquímica y Proteómica Analíticas, Instituto Pasteur Montevideo
Departamento de Neurofisiología Celular y Molecular, Instituto de Investigaciones Biológicas Clemente Estable.

Profesor Responsable de la asignatura:

Flavio Pazos Obregón

Doctor en Ciencias Biológicas por PEDECIBA.
Investigador Grado 3 de PEDECIBA
Investigador Adjunto, Unidad de Bioquímica y Proteómica Analíticas, Instituto Pasteur Montevideo
Investigador Nivel III, Departamento de Neurofisiología Celular y Molecular, Instituto de Investigaciones Biológicas Clemente Estable.

Otros docentes:

Sofía Zeballos

Licenciada en Bioquímica por la UDELAR, Estudiante de la Maestría en Bioinformática PEDECIBA
Laboratorio de Biología de Gusanos, Instituto Pasteur Montevideo
Departamento de Neurofisiología Celular y Molecular, Instituto de Investigaciones

Javiera Quiroz

Ingeniera Biomédica por la Universidad de Concepción, Chile. Estudiante de la Maestría en Bioinformática e PEDECIBA
Departamento de Neurofisiología Celular y Molecular, Instituto de Investigaciones

Docentes invitados¹:

Martín Palazzo

Doctor en Ingeniería por la Université de Technologie de Troyes, Francia
Investigador Asociado de la Universidad de San Andrés, Argentina
Machine Learning research leader en Stamm Biotech, Buenos Aires, Argentina

Gabriel Illanes

Magíster en Ingeniería Matemática, candidato a Doctor en Matemáticas UDELAR
Científico de datos en IDATHA, Montevideo, Uruguay

Carga horaria

- Horas clase (teórico): 9
- Horas clase (práctico): 12
- Horas consulta: 4
- Horas evaluación: 1

◦ **Subtotal horas docencia directa: 26**

- Horas estudio: 18
- Horas resolución ejercicios/prácticos: 6
- Horas proyecto final/monografía: 25

◦ **Subtotal de horas de trabajo domiciliario: 49**

Horas Totales: 75

Nº de Créditos: 5

Público objetivo:

El curso está dirigido a estudiantes de posgrado y estudiantes avanzados de grado de todas las ramas de las ciencias de la vida.

Cupos:

Mínimo: 4

Máximo: 15

Objetivos:

El objetivo del curso es introducir a los estudiantes en el uso de herramientas de aprendizaje automático para el abordaje de problemas biológicos. El curso se compone de clases teóricas y laboratorios prácticos. Las clases teóricas introducirán a los estudiantes en diversos aspectos relativos al aprendizaje automático, haciendo énfasis en la comprensión conceptual de éstos métodos. En las clases prácticas se trabajará en torno a artículos científicos que aborden preguntas biológicas implementando algunos de los métodos vistos en los teóricos. Se trata de un curso introductorio, dirigido a estudiantes sin formación específica ni en estadística ni en programación.

Conocimientos previos exigidos:

- Conocimiento básicos de estadística y álgebra lineal, como los cubiertos por los cursos de grado de Bioestadística y Matemática de las licenciaturas de Biología o Bioquímica de Facultad de Ciencias.

Conocimientos previos recomendados:

- Si bien no es imprescindible, para un mejor aprovechamiento del curso es muy recomendable cierta familiaridad con algún lenguaje de programación (por ejemplo, R o Python).

Metodología de enseñanza:

Virtual, sincrónico, con control de asistencia.

Las clases teóricas consistirán en presentaciones a cargo del docente responsable y de los docentes invitados. En las clases prácticas se trabajará en la plataforma Google Colab, utilizando scripts de Python previamente preparados, con los cuales se reproducirán análisis seleccionados de la bibliografía reciente.

Asistencia:

Asistencia obligatoria al 80% de las clases teóricas y practicas.

Forma de evaluación:

La ganancia del curso se obtiene asistiendo al menos al 80% de las clases teóricas y prácticas y tendrá una vigencia de un año.

Para aprobar el curso se deberá abordar un problema biológico utilizando algunas de las herramientas trabajadas y elaborar un informe escrito que será defendido en una instancia oral junto al equipo docente. El problema a abordar y los datos a utilizar serán definidos en cada caso entre el estudiante y los docentes responsables del curso, buscando vincularlos a proyectos propios de cada estudiante.

El equipo docente evaluará tanto el informe escrito como su presentación oral, cuyo fecha será acordada en cada caso.

Temario:

Semana 1 - Introducción.

Conceptos generales sobre aprendizaje automático. Ejemplos de caso en el abordaje de problemas biológicos.

Semanas 2 y 3 - Aprendizaje no supervisado.

Reducción de dimensionalidad, embeddings y selección de variables. Análisis de componentes principales, tSNE. Autoencoders. Clustering. Algoritmos: K medias, clustering jerárquico. Métricas de evaluación

Semanas 4 y 5 - Aprendizaje supervisado.

Clasificación y regresión. Entrenamiento y evaluación. Validación cruzada, búsqueda en grilla. Métricas de evaluación. Algoritmos: Random Forests, SVM. Perceptrón multicapa y redes neuronales.

Semana 6 - Aprendizaje profundo.

Redes convolucionales. recurrentes y generativas.

Semana 7 - Definición de temas y datos a utilizar para cada informe

Bibliografía:

Introduction to Machine Learning with Python

by Andreas C. Müller and Sarah Guido

O'Reilly Media, Inc. ISBN: 978-1-449-36941-5.

Fecha de inicio y finalización:

Inicio: miércoles 1 de noviembre

Finalización: viernes 13 de diciembre

Institución, Salón y Horario:

Clases virtuales, sincrónicas y con control de asistencia.

Teóricos; miércoles de 15:00 a 16.30 hs

Prácticos; viernes de 15:00 a 17:00 hs
