



Universidad Nacional de Luján
Departamento de
Ciencias Básicas



DISPOSICION CONSEJO DIRECTIVO DEPARTAMENTAL DE CIENCIAS BÁSICAS DISPCD-CB : 243 / 2025

LUJAN, 8 DE JULIO DE 2025

VISTO: El programa de la asignatura Análisis y preprocesamiento de datos biológicos (18926) para la carrera Especialización en Bioinformática; y

CONSIDERANDO:

Que tomó intervención la Comisión Académica de la Carrera.

Que se ha tratado y aprobado por el Consejo Directivo Departamental de Ciencias Básicas en su Sesión Ordinaria del día 3 de julio de 2025.

Por ello,

EL CONSEJO DIRECTIVO DEPARTAMENTAL
DE CIENCIAS BÁSICAS

D I S P O N E :

ARTÍCULO 1°.- APROBAR el programa de la asignatura Análisis y preprocesamiento de datos biológicos (18926) para la carrera Especialización en Bioinformática, que como anexo I forma parte de la presente Disposición.

ARTÍCULO 2°.- ESTABLECER que el mismo tendrá vigencia para los años 2025-2027.-

ARTÍCULO 3°.- Regístrese, comuníquese, cumplido, archívese.-

Lic. Ariel H. REAL - Secretario Académico - Departamento de Ciencias Básicas

Lic. Emma L. FERRERO - Directora Decana - Departamento de Ciencias Básicas

DENOMINACIÓN DE LA ACTIVIDAD: 18926 - **Análisis y preprocesamiento de datos biológicos**

TIPO DE ACTIVIDAD ACADÉMICA: Curso.

CARRERA/S: Especialización en Bioinformática

PLAN DE ESTUDIOS: Plan de Estudios: 76.01 (Resolución HCS N° 711/22)

DOCENTE/S RESPONSABLE/S: Alberto Penas Steinhardt, Doctor UBA; Pablo Reeb, PhD Michigan State University .

EQUIPO DOCENTE: Alberto Penas Steinhardt, Doctor UBA; Pablo Reeb, PhD Michigan State University .

ACTIVIDADES CORRELATIVAS PRECEDENTES:

PARA CURSAR: TRAMO NIVELATORIO

PARA APROBAR: TRAMO NIVELATORIO

MODALIDAD DE DICTADO: (marque con una X donde corresponda).

PRESENCIAL: (Completar en el caso que corresponda según la Resolución

RESHCS-LUJ: 0000283-23: Alternada – Híbrida – Mixta.

DISTANCIA: .

CARGA HORARIA TOTAL: 48 horas

DISTRIBUCIÓN INTERNA DE LA CARGA HORARIA:

TEÓRICO 42% y 20 horas - PRÁCTICO 58% y 28 horas].

Para los espacios curriculares que desarrollen actividades a distancia se debe aclarar la modalidad en la distribución de la carga horaria total (porcentajes de actividades presenciales, a distancia y las estrategias de hibridación: estrategia de "Alternancia", estrategia "Híbrida" o estrategia "Mixta"). En función de las estrategias de hibridación adoptada se debe aclarar la distribución del tiempo [% ACTIVIDADES ASINCRÓNICAS Y SINCRÓNICAS].

PERÍODO DE VIGENCIA DEL PRESENTE PROGRAMA: 2025-2027.

CONTENIDOS MÍNIMOS O DESCRIPTORES:

Transformaciones de variables. Preprocesamiento: normalización de variables, discretización, manejo de ruido, datos faltantes. Reducción de dimensionalidad (Análisis de componentes principales y Análisis Factorial). Estudio de relaciones entre variables. Búsqueda en bases de datos de secuencias expresadas. Tipos de bases de datos biológicas. Referencias cruzadas con otras bases de datos. Bases de datos de secuencias. Principales bases de datos.

FUNDAMENTACIÓN, OBJETIVOS, COMPETENCIAS:

Que los estudiantes logren:

Construir modelos empíricos, con herramientas estadísticas, para fenómenos gaussianos, binomiales, Poisson y de sobrevida, manejo de ruido, datos faltantes, reducción de dimensionalidad con la finalidad de explicar dichos fenómenos o predecir su comportamiento.

Diseñar y analizar experimentos biológicos y computacionales.

Conocer y descargar información de las principales bases de datos biológicas (bases de datos de genes, genomas, secuencias expresadas, proteínas, etc). Obtener referencias cruzadas entre dichas bases de datos.

Aplicar los conocimientos antes mencionados a aplicaciones novedosas.

CONTENIDOS:

Unidad 1: El rol de la Estadística en la Bioinformática: el pensamiento estadístico. Recolección de datos: Principios básicos, estudios retrospectivos, observacionales, diseño de experimentos, observación de procesos a lo largo del tiempo.

Unidad 2: Diseño y análisis de experimentos: Preprocesamiento. Análisis de datos: filtrado de datos, normalización y transformación de variables, discretización, manejo de ruido, datos faltantes. Selección del diseño experimental, ajuste del modelo estadístico adecuado. Análisis de sensibilidad. Validación.

Unidad 3: Reducción de dimensionalidad (Análisis de componentes principales y Análisis Factorial). Estudio de relaciones entre variables.

Unidad 4: Plataformas utilizadas para análisis genómicos y transcriptómicos. Tipos de diseños experimentales: comparaciones directas e indirectas, diseños en bloque, experimentos longitudinales, diseños complejos. Réplicas.

Unidad 5: Tipos de bases de datos biológicas. Búsqueda en bases de datos de secuencias expresadas. Referencias cruzadas con otras bases de datos. Bases de datos de secuencias.

METODOLOGÍA:

Se realizarán prácticas en laboratorios de computación de acuerdo a los requerimientos de cada una, de manera de ejercitar e integrar los contenidos abordados durante las clases teóricas. Cada unidad tiene asociada un trabajo práctico con problemas a resolver; en las horas destinadas a actividades prácticas se resolverán algunos ejercicios testigo y el resto deberán ser resueltos por el estudiante en forma grupal.

La resolución de los trabajos se realizará a través del uso servidores/clusters instalados en la universidad en caso de ser requerido. Esta herramienta permitirá el trabajo colaborativo remoto entre todos los estudiantes del curso. Tiene un doble objetivo: comprender y aplicar los contenidos de cada unidad, al mismo tiempo que se aprende a trabajar en forma remota con una máquina virtual, la cual contendrá todo el software necesario para desarrollar cada unidad del programa de estudios.

TRABAJOS PRÁCTICOS:

Trabajo Práctico 1: Preprocesamiento y análisis exploratorio de datos biológicos

Descripción general: Este trabajo consiste en tomar un conjunto de datos biológicos crudos y realizar un análisis exploratorio completo. Los estudiantes deberán aplicar técnicas de preprocesamiento (filtrado, normalización, manejo de datos faltantes) y explorar relaciones entre variables mediante estadísticas descriptivas y visualizaciones. Se espera que seleccionen un diseño experimental básico y justifiquen su elección.

Trabajo Práctico 2: Reducción de dimensionalidad y análisis de bases de datos genómicas

Descripción general: Los estudiantes trabajarán con un dataset biológico de alta dimensionalidad para aplicar técnicas de reducción de dimensionalidad (Análisis de Componentes Principales o Análisis Factorial). Además, realizarán búsquedas en bases de datos biológicas (ej. NCBI, Ensembl) y compararán resultados con referencias cruzadas, evaluando la consistencia de un diseño experimental complejo (ej. longitudinal o en bloque).

REQUISITOS DE APROBACIÓN Y CRITERIOS DE CALIFICACIÓN:

A lo largo de todo el curso se deberán entregar y aprobar, en las fechas previstas, un trabajo práctico por cada unidad del programa del curso, los cuales se utilizarán para realizar una evaluación formativa del estudiante de carácter continuo.

A su vez, al final del curso deberán aprobar un examen integrador de carácter individual.

BIBLIOGRAFÍA:

Fox, J. (2015). Applied regression analysis and generalized linear models. Sage Publications.

Santner, T. J., Williams, B. J., & Notz, W. I. The Design and Analysis of Computer Experiments (Springer Series in Statistics) 2nd edición. 2018 Edition

Lawson, J. (2014). Design and Analysis of Experiments with R. Chapman and Hall/CRC.

Hoja de firmas