



Facultad de
Ciencias
UNAM



Diplomado Métodos estadísticos y matemáticos para la ciencia de datos

Coordinador: Dr. Gonzalo Pérez de la Cruz. Profesor Asociado C de Tiempo Completo. Departamento de Matemáticas, Facultad de Ciencias, UNAM.

Cuerpo docente

1. Dra. Yuriria Cortés Poza. Investigadora Asociada C de Tiempo Completo. Unidad Académica del IIMAS en el estado de Yucatán, UNAM
2. M. en C. Mario Alejandro Molina Palma. Posgrado en Ciencias Matemáticas y adscrito al Departamento de Probabilidad y Estadística, IIMAS, UNAM
3. M. en E.I.O. José Salvador Zamora Muñoz. Profesor de asignatura, Departamento de Matemáticas, Facultad de Ciencias, UNAM
4. M. en C. David Chaffrey Moreno Fernández. Profesor de asignatura, Departamento de Matemáticas, Facultad de Ciencias, UNAM
5. Dr. Ricardo Ramírez Aldana. Profesor de asignatura, Departamento de Matemáticas, Facultad de Ciencias, UNAM
6. Dr. Gonzalo Pérez de la Cruz. Profesor Asociado C de Tiempo Completo. Departamento de Matemáticas, Facultad de Ciencias, UNAM

TEMARIO

Módulos	Subtemas	No. de horas	Ponente(s)	Fechas
I. Preprocesamiento y técnicas básicas de visualización de datos	<p>I.1. Introducción a Python y R para el manejo y preprocesamiento de datos</p> <p>I.1.1. Manejo básico de datos: Importar, filtrar, ordenar y mezclar bases de datos, así como creación de nuevas variables. Uso de los paquetes numpy, pandas y tidyverse.</p> <p>I.1.2. Programación básica: definición de funciones, uso de ciclos y estructuras condicionales.</p> <p>I.1.3. Uso de algoritmos y estructuras de datos eficientes</p> <p>I.2. Técnicas básicas de visualización y exploración de datos</p> <p>I.2.1. Tipos de variables y estadísticas descriptivas básicas</p> <p>I.2.2. Creación de gráficas básicas: diagramas de dispersión, mapas de calor de correlaciones, etc. Uso de los paquetes matplotlib y ggplot2</p> <p>I.2.3. Transformaciones básicas a las variables: normalización, categorización, etc.</p> <p>I.2.4. Identificación de outliers y manejo básico de datos perdidos.</p>	40	David Chaffrey Moreno Fernández	4 oct-9 nov, 2024
II. Métodos de aprendizaje no supervisado para explorar, describir y visualizar los datos.	<p>II.1. Aprendizaje supervisado vs no supervisado. Supuestos básicos y ejemplos de aplicaciones.</p> <p>II.2. Análisis de conglomerados.</p> <p>II.2.1. Métodos de partición</p> <ol style="list-style-type: none"> K-Medias Hierarchical K-means clustering Fuzzy clustering <p>II.2.2. Métodos jerárquicos.</p> <ol style="list-style-type: none"> Métodos aglomerativos y divisivos 	40	José Salvador Zamora Muñoz	15 nov-14 dic, 2024

	<ul style="list-style-type: none"> ii. Disimilaridades entre observaciones y entre conglomerados <p>II.3. Métodos de reducción de dimensionalidad</p> <p>II.3.1. Análisis de componentes principales</p> <ul style="list-style-type: none"> i. Construcción e interpretación de los componentes principales ii. Biplot <p>II.3.2. Análisis factorial exploratorio.</p> <ul style="list-style-type: none"> i. Construcción e Interpretación de los factores ii. Herramientas para la determinación del número de factores <p>II.3.3. Escalamiento multidimensional</p> <ul style="list-style-type: none"> i. Escalamiento métrico ii. Escalamiento no métrico <p>II.3.4. T-distributed Stochastic Neighbor Embedding (t-SNE)</p>			
<p>III. Introducción a los modelos lineales generalizados (glm) con el objetivo de estimación e inferencia. Parte 1.</p>	<p>III.1. El modelo de regresión lineal múltiple</p> <ul style="list-style-type: none"> III.1.1. Estimación y pruebas de hipótesis. III.1.2. Diagnóstico de los supuestos. III.1.3. Transformaciones de Box-Cox y Box-Tidwell. III.1.4. Regresión polinomial. <p>III.2. Técnicas para la selección de modelos.</p> <ul style="list-style-type: none"> III.2.1. Criterios AIC y BIC III.2.2. Métodos stepwise III.2.3. Métodos de regularización (lasso y ridge) <p>III.3. Ejemplos de problemas tipo ANOVA y ANCOVA</p> <ul style="list-style-type: none"> III.3.1. Uso de variables explicativas cualitativas III.3.2. Relación con el diseño experimental y el objetivo de causalidad 	32	Ricardo Ramírez Aldana	10 ene-1 feb, 2025
<p>IV. Introducción a los modelos lineales generalizados</p>	<p>IV.1. Modelos lineales generalizados para variables respuesta cualitativas y conteos</p>	32	Gonzalo Pérez de la Cruz	7 feb-1 mar, 2025

<p>(glm) con el objetivo de estimación e inferencia. Parte 2.</p>	<p>IV.1.1. Uso de la regresión logística, multinomial, Poisson, binomial negativa y modelos loglineales</p> <p>IV.1.2. Estimación y diagnóstico de los supuestos.</p> <p>IV.1.3. Criterios AIC y BIC, así como técnicas para selección de modelos: métodos stepwise y métodos de regularización (lasso, ridge).</p> <p>IV.2. Tablas de contingencia: Tipos de independencia e Introducción a modelos gráficos probabilísticos (caso discreto)</p>			
<p>V. Métodos clásicos para el objetivo de predicción.</p>	<p>V.1. El problema de predicción en nuevas observaciones</p> <p>V.1.1. Definiciones básicas</p> <ol style="list-style-type: none"> i. Reglas o algoritmos de predicción ii. Parámetros e hiperparámetros iii. Entrenamiento (estimación o definición de parámetros e hiperparámetros, selección de modelos) vs evaluación de la regla (cálculo del poder predictivo). <p>V.1.2. Métodos para calcular el poder predictivo</p> <ol style="list-style-type: none"> i. Métricas de desempeño <ol style="list-style-type: none"> a. Caso continuo: Error Medio Absoluto (MAE); Porcentaje de Error Medio Absoluto (MAPE); Error cuadrático medio (MSE) b. Caso discreto: Accuracy (Exactitud); Tasa de error de clasificación; Sensibilidad; Especificidad; Curva ROC ii. Conjuntos de entrenamiento y prueba <ol style="list-style-type: none"> a. Ajuste o definición de la regla con los datos de entrenamiento b. Evaluación de la regla con el conjunto de prueba iii. Métodos de remuestreo: Repeated holdout y Cross-validation <p>V.2. Métodos clásicos para definir reglas de predicción de valores continuos</p> <p>V.2.1. Modelos de regresión lineal múltiple y otros modelos lineales generalizados</p>	<p>32</p>	<p>Gonzalo Pérez de la Cruz</p>	<p>7-29 mar, 2025</p>

	<p>V.2.2. Árboles de regresión</p> <p>V.2.3. Regresión tipo k-vecinos más cercanos</p> <p>V.2.4. Métodos para datos de series de tiempo</p> <p>V.3. Métodos clásicos para definir reglas de clasificación</p> <p>V.3.1. Regresión logística y multinomial</p> <p>V.3.2. Análisis de discriminante</p> <p>V.3.3. K vecinos más cercanos</p> <p>V.3.4. Naive bayes classifier</p> <p>V.3.5. Árboles de clasificación</p>			
VI. Métodos modernos para el objetivo de predicción.	<p>VI.1. Calibración de hiperparámetros (tuneo)</p> <p>VI.2. Métodos</p> <p>VI.2.1. Métodos de ensamble tipo bootstrap: bagging, random forests y boosting</p> <p>VI.2.2. Máquinas de Soporte Vectorial</p> <p>VI.2.3. Introducción a las Redes Neuronales y al aprendizaje profundo</p>	32	Ricardo Ramírez Aldana	4 abr-3 may, 2025
VII. Optimización en aprendizaje de máquina	<p>VII.1. Introducción a la optimización</p> <p>VII.1.1. Modelos y funciones de pérdida</p> <p>VII.1.2. Panorama de optimización</p> <p>VII.1.3. Búsqueda en rejilla (grid search)</p> <p>VII.1.4. Métodos de primer orden</p> <p>VII.1.5. Métodos de segundo orden</p> <p>VII.2. Descenso del gradiente y sus variaciones</p> <p>VII.2.1. Descenso del gradiente</p> <p>VII.2.2. Descenso del gradiente estocástico</p> <p>VII.2.3. Variaciones adaptativas</p> <p>VII.2.4. Variaciones aleatorizadas</p> <p>VII.3. Optimización convexa y no-convexa</p> <p>VII.3.1. Optimización de funciones convexas</p> <p>VII.3.2. Optimización de funciones no-convexas</p>	32	Mario Alejandro Molina Palma y Yuriria Cortés Poza	9 may-6 jun, 2025

	<ul style="list-style-type: none">VII.3.3. Efecto de la inicialización en la optimizaciónVII.3.4. Efecto de la tasa de aprendizaje en la optimizaciónVII.3.5. Efecto de otros hiperparámetros en la optimización <p>VII.4. Otros paradigmas de optimización</p> <ul style="list-style-type: none">VII.4.1. Aprendizaje en líneaVII.4.2. Aprendizaje por refuerzo			
--	---	--	--	--