



Universidad
Nacional
de Córdoba



FAMAF
Facultad de Matemática,
Astronomía, Física y
Computación

EX-2024-00149385- -UNC-ME#FAMAF

PROGRAMA DE ASIGNATURA	
ASIGNATURA: Ciencia de Datos	AÑO: 2024
CARACTER: Obligatoria	UBICACIÓN EN LA CARRERA: 4° año 1° cuatrimestre
CARRERA: Licenciatura en Matemática Aplicada	
REGIMEN: Cuatrimestral	CARGA HORARIA: 120 Horas.

FUNDAMENTACIÓN Y OBJETIVOS

Este curso introduce al/a la estudiante a los tópicos generales de Ciencia de datos. El curso comienza con una discusión sobre las diferencias entre el aprendizaje automático y el análisis multivariado clásico. Se discutirán temas centrales del área como son reducción de dimensión, creación de clasificadores a partir de definición de hipótesis minimales, riesgo y error y métodos de agrupamiento basados en métricas. Se estudiarán también errores y medidas de desempeño.

CONTENIDO

Unidad 1 : Manejo de datos y visualización

Cómo dar a una computadora la habilidad de aprender de los datos. Tres formas de aprendizaje por computadora. Notación y terminología técnica. Uso de Python. Pandas, Plotly, Seaborn, Matplotlib.

Unidad 2: Hipótesis determinísticas

Aprendizaje de conceptos, algoritmos Find S, Eliminación completa (Complete elimination) y Eliminación de candidatos (Candidate elimination). Árboles de decisión, algoritmo ID3. Bagging. Boosting. Algoritmos Random Forests y Ada Boost.

Unidad 3: Hipótesis estadísticas

Discriminante de Bayes, caso gaussiano, Discriminante de Fisher. Funciones discriminantes multicaso. Redes de creencias (Belief networks), Bayesiano ingenuo (Naive Bayes). Estimación paramétrica y bayesiana. Mezcla de gaussianas. Algoritmo Esperanza-Maximización (Expectation Maximization).

Unidad 4: PCA, LDA CC y otras A's

Análisis de componentes principales y discriminantes, correlación canónica, análisis de componentes independientes y t SNE, Stochastic Embedding.

Unidad 5: Métodos lineales

Métodos para problemas con clases linealmente separables. Perceptron. Algoritmos de optimización para cálculo de hiperplanos. Regresión logística, Máquinas de Vectores de Soporte. Algoritmos Kernel Support Vector Machine. Esquemas de discriminación multiclase.

Unidad 6: Aprendizaje no supervisado

Algoritmo apriori. K-medias, métodos jerárquicos, mezcla de Gaussianas. Algoritmos Mean Shift , DBscan, Optics, y Birch.

Métodos de selección del número de clusters.

Visualización usando t-SNE

BIBLIOGRAFÍA

BIBLIOGRAFÍA BÁSICA

Pattern Classification. R. Duda, P. Hart y D.Stork, Wiley 2006

Python machine learning. SI Rashka. Packt 2016.

EX-2024-00149385- -UNC-ME#FAMAF

Machine Learning Tom M. Mitchell McGraw-Hill, 1997

EVALUACIÓN

FORMAS DE EVALUACIÓN

Deberán entregar todos los trabajos prácticos derivados de la ejercitación general.
Tendrán dos instancias de evaluación parcial en el aula.

REGULARIDAD

1. cumplir un mínimo de 70% de asistencia a clases teóricas, prácticas y de laboratorio.
2. aprobar al menos dos evaluaciones parciales o sus correspondientes recuperatorios.
3. aprobar al menos el 60% de los Trabajos Prácticos o de Laboratorio.

PROMOCIÓN

1. cumplir un mínimo de 80% de asistencia a clases teóricas, prácticas, o de laboratorio.
2. aprobar todas las evaluaciones parciales con una nota no menor a 6 (seis), y obteniendo un promedio no menor a 7 (siete).
3. aprobar todos los Trabajos Prácticos o de Laboratorio.
4. Aprobar el coloquio con nota mayor a 7 (siete). La nota final será un promedio entre los parciales y la nota del coloquio.