



**CONAHCYT**

CONSEJO NACIONAL DE HUMANIDADES  
CIENCIAS Y TECNOLOGÍAS



**CIMAT**



**Nombre de la línea terminal:** Ciencia de datos

Descripción

Perfil de ingreso

Perfiles afines a esta línea terminal

Perfil de egreso

## Materias para EMED

**Módulo 1 Fundamentos de programación:**

Temario

### 1. Fundamentos de Python

**Objetivo:** El alumno aprenderá los comandos básicos de Python y habrá instalado software en su sistema operativo de preferencia

#### 1.1 Introducción

##### 1.1.1 Preliminares

##### 1.1.2 Conceptos básicos del lenguaje Python, IPython y Jupyter Notebooks

##### 1.1.3 Estructura de datos, funciones y archivos integrados

### 2. Python aplicado a la ciencia de datos

**Objetivo:** El alumno conocerá las principales funciones de las librerías NumPy, Pandas y Matplotlib

#### 2.1 NumPy

#### 2.2 Pandas

#### 2.3 Matplotlib



**CONAHCYT**

CONSEJO NACIONAL DE HUMANIDADES  
CIENCIAS Y TECNOLOGÍAS



**CIMAT**



### 3. Introducción a base de datos

Objetivo: Que el alumno conozca las bases de datos, así como los elementos básicos asociados a ellas. Además, el alumno será capaz de realizar operaciones sobre las bases de datos para la consulta de información y modificación.

- 3.1 Conceptos de bases de datos
- 3.2 El sistema SQLite
- 3.3 Consulta de elementos
- 3.4 Administración de base de datos

### 4. R aplicado a ciencia de datos

## Módulo 2 Optimización

### 1. Optimización sin restricciones

- 1.1 Nociones de probabilidad
  - 1.1.1 La derivada y su interpretación gráfica
  - 1.1.2 Máximos y mínimos de una función
- 1.2 Optimización
  - 1.2.1 Óptimo local y óptimo global
  - 1.2.2 Condiciones de optimalidad
- 1.3 Optimización numérica
  - 1.3.1 Método del descenso empinado
  - 1.3.2 Método de Newton
  - 1.3.3 Métodos de cuasi-Newton
- 1.4 Ejemplo práctico

### 2. Optimización con restricciones

- 2.1 Introducción
  - 2.1.1 Multiplicadores de Lagrange
  - 2.1.2 Condiciones de Karush-Kuhn-Tucker
  - 2.1.3 Programación lineal
  - 2.1.4 Programación cuadrática

### 3. Optimización combinatoria

- 3.1 Introducción



- 3.2 Programación entera
- 3.3 Teoría de grafos
  - 3.3.1 Camino más corto
  - 3.3.2 Problema del flujo máximo
  - 3.3.3 Problema del flujo de costo mínimo

- 4. Optimización estocástica
  - 4.1 ¿Qué es la programación estocástica?
  - 4.2 Gradiente estocástico
  - 4.3 Búsqueda aleatoria
  - 4.4 Evolución diferencial
  - 4.5 Optimización por enjambre de partículas

## Módulo 3 Métodos multivariados II (Métodos de aprendizaje no supervisado)

- 1. Técnicas de reducción de dimensión [ACP]
  - 1.1 Análisis de componentes principales
    - 1.1.1 Planteamiento
    - 1.1.2 Conceptos básicos
    - 1.1.3 Definición y obtención de los componentes principales
    - 1.1.4 Selección del número de componentes principales
    - 1.1.5 Representación gráfica
    - 1.1.6 Pasos para realizar un análisis de componentes principales
  - 1.2 Análisis de correspondencia
    - 1.2.1 Introducción
    - 1.2.2 Pasos para la aplicación
    - 1.2.3 Análisis de correspondencia múltiple
    - 1.2.4 Resumen
  - 1.3 Regresión por mínimos cuadrados parciales [PLS]
    - 1.3.1 Regresión PLS y su relación con el análisis de correlación canónica
    - 1.3.2 Regresión PLS y su relación con regresión de componentes principales
    - 1.3.3 Motivación de la regresión PLS
    - 1.3.4 Regresión PLS con una única variable respuesta Y
    - 1.3.5 Validación cruzada para determinar el número de componentes adecuados

## 2. Aprendizaje no supervisado



- 2.1 Aprendizaje no supervisado (análisis de conglomerados)
- 2.2 Tipos de análisis de conglomerados
- 2.3 Medidas de proximidad
- 2.4 Métodos Jerárquicos
- 2.5 Métodos no jerárquicos o de partición

## Módulo 4. Métodos de aprendizaje supervisado

### 1. Clasificación

- 1.1 Clasificación óptima
- 1.2 Clasificación Bayesiano óptimo
- 1.3 Clasificación óptima con poblaciones normales
- 1.4 Ejemplo: Estudio sobre cáncer de mama
- 1.5 Estimación de parámetros
- 1.6 Discriminante con K poblaciones
- 1.7 Fisher y LDA

### 2. Regresión logística

- 2.1 el modelo logístico
- 2.2 Máxima verosimilitud
- 2.3 Interpretación de los coeficientes del modelo
- 2.4 Errores estándar
- 2.5 Clasificación con regresión logística
- 2.6 Prueba de hipótesis
- 2.7 Curvas ROC Redes neuronales

### 3. Redes neuronales

- 3.1 El algoritmo perceptrón
- 3.2 Redes neuronales multicapa
  - 3.2.1 Capas y unidades ocultas
  - 3.2.2 Unidades de salida
  - 3.2.3 Funciones de costo
  - 3.2.4 Ajuste
  - 3.2.5 Regularización
  - 3.2.6 Extensión a RN profundas árboles de clasificación y regresión

### 4. Máquinas de soporte vectorial

- 4.1 Hiperplanos separadores
- 4.2 SVM caso separable
- 4.3 SVM caso no separable



**CONAHCYT**

CONSEJO NACIONAL DE HUMANIDADES  
CIENCIAS Y TECNOLOGÍAS



**CIMAT**



- 4.4 Kernels: Fronteras no lineales
- 4.5 Validación cruzada
- 4.6 El caso multiclase
  
- 5. Árboles de clasificación y regresión
  - 5.1 Construcción de un árbol de clasificación
    - 5.1.1 Estrategias para dividir un nodo
    - 5.1.2 Estrategias para crear un árbol 0
    - 5.1.3 Estrategias para asignar nodos terminales
    - 5.1.4 Selección del árbol óptimo
    - 5.1.5 Importancia de características
  
- 6. Métodos de ensamble
  - 6.1 El concepto de varianza, sesgo y complejidad del modelo
    - 6.1.1 Selección de modelos mediante validación cruzada
  - 6.2 Bagging
    - 6.2.1 Evaluación out-of-bag
    - 6.2.2 Bagging con árboles de clasificación
  - 6.3 Boosting
    - 6.3.1 AdaBoost
    - 6.3.2 La función de costo de AdaBoost
    - 6.3.3 Modelos aditivos
    - 6.3.4 AdaBoost ajusta un modelo aditivo
    - 6.3.5 AdaBoost multiclase
  - 6.4 Random Forest
  - 6.5 Una comparativa de los métodos de ensamble