

**NOMBRE DE LA ASIGNATURA O UNIDAD DE APRENDIZAJE.**

Aprendizaje de Máquina

**CICLO III****CLAVE DE LA ASIGNATURA****OBJETIVO(S) GENERAL(ES) DE LA ASIGNATURA**

Proporcionar una base sólida en los fundamentos teóricos y prácticos del aprendizaje automático clásico, enfatizando en modelos interpretables, técnicas de evaluación y aplicaciones en problemas del mundo real,

**TEMAS Y SUBTEMAS**

## 1. Análisis Discriminante Lineal

## 1.1 Introducción

## 1.2 Análisis Discriminante Gaussiano

## 1.2.1 Fronteras de Decisión Cuadráticas

## 1.2.2 Fronteras de Decisión Lineales

## 1.2.3 Conexión entre LDA y Regresión Logística

## 1.2.4 Ajuste del Modelo

## 1.2.5 Clasificador del Centroide Más Cercano

## 1.2.6 Análisis Discriminante Lineal de Fisher\*

## 1.3 Clasificadores de Naive Bayes

## 1.3.1 Modelos de Ejemplo

## 1.3.2 Ajuste del Modelo

## 1.3.3 Naive Bayes Bayesiano

## 1.3.4 Conexión entre Naive Bayes y Regresión Logística

## 1.4 Clasificadores Generativos vs Discriminativos

## 1.4.1 Ventajas de los Clasificadores Discriminativos

## 1.4.2 Ventajas de los Clasificadores Generativos

## 1.4.3 Manejo de Características Faltantes

## 2. Regresión Logística

## 2.1 Introducción

- 2.2 Regresión Logística Binaria
  - 2.2.1 Clasificadores Lineales
  - 2.2.2 Clasificadores No Lineales
  - 2.2.3 Estimación de Máxima Verosimilitud
  - 2.2.4 Descenso de Gradiente Estocástico
  - 2.2.5 Algoritmo del Perceptrón
  - 2.2.6 Mínimos Cuadrados Reponderados Iterativos
  - 2.2.7 Estimación MAP
  - 2.2.8 Estandarización
- 2.3 Regresión Logística Multinomial
  - 2.3.1 Clasificadores Lineales y No Lineales
  - 2.3.2 Estimación de Máxima Verosimilitud
  - 2.3.3 Optimización Basada en Gradiente
  - 2.3.4 Optimización por Cotas
  - 2.3.5 Estimación MAP
  - 2.3.6 Clasificadores de Máxima Entropía
  - 2.3.7 Clasificación Jerárquica
  - 2.3.8 Manejo de un Gran Número de Clases
- 2.4 Regresión Logística Robusta
  - 2.4.1 Modelo de Mezcla para la Verosimilitud
  - 2.4.2 Pérdida Bi-Templada
- 2.5 Regresión Logística Bayesiana
  - 2.5.1 Aproximación de Laplace
  - 2.5.2 Aproximación de la Predictiva Posterior
- 3. Regresión Lineal
  - 3.1 Introducción
  - 3.2 Regresión Lineal por Mínimos Cuadrados
    - 3.2.1 Terminología
    - 3.2.2 Estimación por Mínimos Cuadrados
    - 3.2.3 Otros Enfoques para Calcular el MLE
    - 3.2.4 Medida de la Calidad del Ajuste
  - 3.3 Regresión Ridge
    - 3.3.1 Cálculo de la Estimación MAP
    - 3.3.2 Conexión entre Ridge y PCA
    - 3.3.3 Elección de la Intensidad del Regularizador
  - 3.4 Regresión Lasso
    - 3.4.1 Estimación MAP con Prior de Laplace (Regularización  $\ell_1$ )
    - 3.4.2 ¿Por Qué la Regularización  $\ell_1$  Produce Soluciones Dispersas?
    - 3.4.3 Umbral Fuerte vs Suave
    - 3.4.4 Ruta de Regularización
  - 3.4.5 Comparación de Mínimos Cuadrados, Lasso, Ridge y Selección de Subconjuntos
  - 3.4.6 Consistencia en la Selección de Variables

- 3.4.7 Lasso por Grupos
- 3.4.8 Red Elástica (Combinación de Ridge y Lasso)
- 3.4.9 Algoritmos de Optimización
- 3.5 Splines de Regresión\*
  - 3.5.1 Funciones Base B-Spline
  - 3.5.2 Ajuste de un Modelo Lineal con Base de Splines
  - 3.5.3 Splines Suavizados
  - 3.5.4 Modelos Aditivos Generalizados
- 3.6 Regresión Lineal Robusta\*
  - 3.6.1 Verosimilitud de Laplace
  - 3.6.2 Verosimilitud de Student-t
  - 3.6.3 Pérdida de Huber
  - 3.6.4 RANSAC
- 3.7 Regresión Lineal Bayesiana
  - 3.7.1 Priori
  - 3.7.2 Posteriori
  - 3.7.3 Cálculo de la Predictiva Posterior
  - 3.7.4 Ventaja del Centrando
  - 3.7.5 Manejo de Multicolinealidad
  - 3.7.6 Determinación Automática de Relevancia (ARD)
- 4. Modelos Lineales Generalizados (GLMs)
  - 4.1 Introducción
  - 4.2 Ejemplos
    - 4.2.1 Regresión Lineal
    - 4.2.2 Regresión Binomial
    - 4.2.3 Regresión de Poisson
  - 4.3 GLMs con Funciones de Enlace No Canónicas
  - 4.4 Estimación de Máxima Verosimilitud
  - 4.5 Ejemplo Práctico: Predicción de Reclamaciones de Seguros
- 5. Redes Neuronales para Datos Estructurados
  - 5.1 Introducción
  - 5.2 Perceptrones Multicapa (MLPs)
    - 5.2.1 El Problema XOR
    - 5.2.2 MLPs Diferenciables
    - 5.2.3 Funciones de Activación
    - 5.2.4 Modelos de Ejemplo
    - 5.2.5 La Importancia de la Profundidad
    - 5.2.6 La “Revolución del Aprendizaje Profundo”
    - 5.2.7 Conexiones con la Biología
  - 5.3 Retropropagación
    - 5.3.1 Diferenciación en Modo Directo vs Reverso
    - 5.3.2 Diferenciación en Modo Reverso para MLPs

- 5.3.3 Producto Vector-Jacobiano para Capas Comunes
- 5.3.4 Grafos de Cómputo
- 5.4 Entrenamiento de Redes Neuronales
  - 5.4.1 Ajuste de la Tasa de Aprendizaje
  - 5.4.2 Gradientes Desvanecientes y Explosivos
  - 5.4.3 Funciones de Activación No Saturantes
  - 5.4.4 Conexiones Residuales
  - 5.4.5 Inicialización de Parámetros
  - 5.4.6 Entrenamiento Paralelo
- 5.5 Regularización
  - 5.5.1 Detención Temprana
  - 5.5.2 Decaimiento de Pesos
  - 5.5.3 DNNs Dispersas
  - 5.5.4 Dropout
  - 5.5.5 Redes Neuronales Bayesianas
  - 5.5.6 Efectos de Regularización del Descenso de Gradiente (Estocástico)\*
- 5.6 Otros Tipos de Redes Feedforward\*
  - 5.6.1 Redes con Funciones Radiales de Base
  - 5.6.2 Mezclas de Expertos
- 6. Métodos Basados en Ejemplares
  - 6.1 Clasificación por los K Vecinos Más Cercanos (KNN)
    - 6.1.1 Ejemplo
    - 6.1.2 La Maldición de la Dimensionalidad
    - 6.1.3 Reducción de los Requisitos de Velocidad y Memoria
    - 6.1.4 Reconocimiento de Conjunto Abierto
  - 6.2 Aprendizaje de Métricas de Distancia
    - 6.2.1 Métodos Lineales y Convexos
    - 6.2.2 Aprendizaje Profundo de Métricas (Deep Metric Learning)
    - 6.2.3 Pérdidas para Clasificación
    - 6.2.4 Pérdidas para Ordenamiento (Ranking)
    - 6.2.5 Aceleración de la Optimización con Pérdidas de Ordenamiento
    - 6.2.6 Otros Trucos de Entrenamiento para DML
  - 6.3 Estimación de Densidad con Núcleos (KDE)
    - 6.3.1 Núcleos de Densidad
    - 6.3.2 Estimador de Densidad por Ventana de Parzen
    - 6.3.3 Cómo Elegir el Parámetro de Ancho de Banda
    - 6.3.4 De KDE a Clasificación por KNN
    - 6.3.5 Regresión con Núcleos
- 7. Métodos de Núcleo (Kernel Methods) \*
- 7.1 Núcleos de Mercer
  - 7.1.1 Teorema de Mercer

- 7.1.2 Algunos Núcleos Populares de Mercer
- 7.2 Procesos Gaussianos
  - 7.2.1 Observaciones sin Ruido
  - 7.2.2 Observaciones con Ruido
  - 7.2.3 Comparación con Regresión con Núcleos
  - 7.2.4 Espacio de Pesos vs Espacio de Funciones
  - 7.2.5 Problemas Numéricos
  - 7.2.6 Estimación del Núcleo
  - 7.2.7 Procesos Gaussianos para Clasificación
  - 7.2.8 Conexiones con el Aprendizaje Profundo
  - 7.2.9 Escalado de GPs a Conjuntos de Datos Grandes
- 7.3 Máquinas de Vectores de Soporte (SVMs)
  - 7.3.1 Clasificadores de Gran Margen
  - 7.3.2 El Problema Dual
  - 7.3.3 Clasificadores de Margen Suave
  - 7.3.4 El Truco del Núcleo
  - 7.3.5 Conversión de Salidas SVM en Probabilidades
  - 7.3.6 Conexión con Regresión Logística
  - 7.3.7 Clasificación Multiclas con SVMs
  - 7.3.8 Cómo Elegir el Regularizador C
  - 7.3.9 Regresión Ridge con Núcleos
  - 7.3.10 SVMs para Regresión
- 7.4 Máquinas de Vectores Dispersos
  - 7.4.1 Máquinas de Vectores Relevantes (RVMs)
  - 7.4.2 Comparación entre Métodos de Núcleo Dispersos y Densos
- 8. Árboles, Bosques, Bagging y Boosting
  - 8.1 Árboles de Clasificación y Regresión (CART)
    - 8.1.1 Definición del Modelo
    - 8.1.2 Ajuste del Modelo
    - 8.1.3 Regularización
    - 8.1.4 Manejo de Características de Entrada Faltantes
    - 8.1.5 Ventajas y Desventajas
  - 8.2 Aprendizaje de Ensambls
    - 8.2.1 Apilamiento (Stacking)
    - 8.2.2 Ensamble no es Promediado Bayesiano de Modelos
  - 8.3 Bagging
  - 8.4 Bosques Aleatorios
  - 8.5 Boosting
    - 8.5.1 Modelado Aditivo por Etapas Hacia Adelante
    - 8.5.2 Pérdida Cuadrática y Boosting de Mínimos Cuadrados
    - 8.5.3 Pérdida Exponencial y AdaBoost
    - 8.5.4 LogitBoost

### 8.5.5 Boosting por Gradiente

## 8.6 Interpretación de Ensamblés de Árboles

### 8.6.1 Importancia de las Características

### 8.6.2 Gráficos de Dependencia Parcial

## ACTIVIDADES DE APRENDIZAJE

### 1. Actividades Teóricas

- Sesiones de lecturas guiadas:
  - Análisis de artículos clásicos
  - Discusión de fundamentos matemáticos.
- Resolución de ejercicios teóricos:
  - Demostraciones de teoremas.

### 2. Actividades Prácticas

Laboratorios de código:

- Implementación de algoritmos desde cero.

### 3. Proyectos Integrador.

### 4. Evaluación y Retroalimentación

- Exámenes teóricos y prácticos:
- Presentaciones orales:

### 5. Actividades de Actualización

- Webinars con expertos:
  - Charlas sobre aplicaciones emergentes

### 6. Habilidades Blandas

- Documentación técnica:
  - Creación de tutoriales en Jupyter Notebooks con Markdown.
- Trabajo en equipo:

## Bibliografía

1. Murphy, Kevin P. (2022). *Probabilistic machine learning: an introduction*. Massachusetts: MIT Press. Chicago Style. Murphy, Kevin P.
2. Abu-Mostafa, Y. S., Magdon-Ismail, M., & Lin, H. (2012). *Learning from data: a short course*. AMLBook.com.
3. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2a ed.). Springer.
4. Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
5. Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). *Foundations of machine learning* (2a ed.). MIT Press.

### **CRITERIOS Y PROCEDIMIENTOS DE EVALUACION Y ACREDITACION**

El curso se evaluará con el promedio de resultados de tres evaluaciones parciales, una evaluación final y la presentación de un proyecto integrador. El curso será acreditado con una calificación promedio de al menos 8.0.