

Reconocimiento de Patrones (MIE.VR01.09)

Dr. Mario-Alberto IBARRA-MANZANO (ibarram@ugto.mx)

Enero del 2026

Días: Martes y Jueves
Horario: 14:00-16:00
Aula: Aula de Posgrado

1 Competencia de la Unidad de Aprendizaje

Diseña, implementa y evalúa sistemas de reconocimiento de patrones mediante el pipeline completo (*datos* \rightarrow *preprocesamiento* \rightarrow *representación/rasgos* \rightarrow *modelo* \rightarrow *validación* \rightarrow *métricas*), utilizando fundamentos estadísticos y de aprendizaje automático. Desarrolla soluciones reproducibles en **R** para problemas de clasificación, regresión y agrupamiento, interpretando resultados con pensamiento crítico, considerando limitaciones experimentales (p.ej., fuga de información), desbalance de clases, y criterios de desempeño acordes al dominio de aplicación. Exhibe actitudes como el trabajo en equipos multidisciplinarios, comunicación científica clara y ética profesional.

La UDA de Reconocimiento de Patrones proporciona una base rigurosa para modelar y analizar datos provenientes de diversos dominios (industria, biomédico, visión, texto, sensores), conectando teoría de decisión, representación de datos y validación estadística. El énfasis en reproducibilidad, buenas prácticas y evaluación robusta permite desarrollar sistemas confiables y transferibles a entornos reales de ingeniería y ciencia de datos.

2 Contenido de la Unidad de Aprendizaje

- Introducción al reconocimiento de patrones y flujo de trabajo en R
 - Definiciones: patrón, clase, etiqueta, característica, variabilidad y ruido
 - Tipos de problemas: clasificación, regresión, clustering y detección de anomalías (panorama)
 - Pipeline completo: *datos* \rightarrow *preprocesamiento* \rightarrow *rasgos* \rightarrow *modelo* \rightarrow *validación* \rightarrow *métricas*

- Reproducibilidad en R: proyectos, semillas, control de versiones (panorama), reportes (Quarto/RMarkdown)
- Diseño experimental: particiones *train/validation/test*, *cross-validation* y prevención de *data leakage*
- Fundamentos probabilísticos y decisión estadística
 - Probabilidad, verosimilitud y estimación (nociones)
 - Teorema de Bayes, priors y posteriors
 - Clasificador Bayesiano óptimo: riesgo esperado, matrices de costo y tasas de error
 - Modelos generativos vs discriminativos (comparación conceptual)
- Preparación de datos y representación de patrones
 - Limpieza de datos: valores faltantes, outliers, escalamiento/normalización, codificación de categóricas
 - Ingeniería de características: transformaciones y construcción de rasgos (principios)
 - Selección de características: filtros, *wrappers* y métodos embebidos (nociones)
 - Reducción de dimensionalidad:
 - * PCA: interpretación geométrica y varianza explicada
 - * Proyecciones no lineales (panorama): t-SNE/UMAP y precauciones
 - Sesgo-varianza y regularización (intuición aplicada)
- Clasificación supervisada clásica
 - kNN: métricas de distancia, elección de k , sensibilidad al escalamiento
 - Regresión logística: frontera lineal, interpretación probabilística y regularización
 - LDA/QDA: supuestos gaussianos, covarianzas y fronteras de decisión
 - SVM: margen, kernels (lineal y RBF), hiperparámetros (C , γ)
 - Calibración de probabilidades (nociones) y selección de umbral
- Ensamblados y métodos basados en árboles
 - Árboles de decisión: impureza, sobreajuste y poda
 - Random Forest: *bagging*, aleatoriedad y relevancia de variables
 - Gradient Boosting: idea general y comparación con *bagging* (panorama)
 - Desbalance de clases: re-muestreo, pesos de clase, ajuste de umbral

- Interpretabilidad (panorama): importancia por permutación, PDP/ICE, SHAP
- Aprendizaje no supervisado y evaluación integral
 - Clustering: k-means, jerárquico y mezcla gaussiana (nociones)
 - Validación de clustering: silhouette, Davies–Bouldin (nociones)
 - Integración del pipeline: comparación de modelos y selección final
 - Reporte reproducible: resultados, discusión crítica, limitaciones y recomendaciones
- Implementación en R (a lo largo del curso)
 - Manejo de datos y EDA: tidyverse (dplyr, tidyr, ggplot2)
 - Modelado reproducible: tidymodels (rsample, recipes, parsnip, workflows, tune, yardstick)
 - Buenas prácticas: separación de datos, resampling, tuning, y control de fuga de información

3 Evaluación

Práctica 1: EDA + diseño experimental (R/Quarto)	15%
Práctica 2: Baselines + métricas (clasificación/regresión)	15%
Evaluación 1: Fundamentos Bayesianos + validación	15%
Práctica 3: Features + PCA + selección (pipeline reproducible)	15%
Evaluación 2: Modelos supervisados (kNN, LR, LDA/QDA, SVM)	15%
Proyecto final (código + reporte reproducible)	15%
Presentación del proyecto (defensa y crítica)	10%

Cada práctica/proyecto será valorado en seis ejes:

- **Desempeño:** 1/6 de la calificación, medido con métricas apropiadas al problema (p. ej., *accuracy*, *F1*, *balanced accuracy*, ROC/PR AUC, RMSE/MAE) y reportadas con intervalos o variación por resampling.
- **Diseño experimental:** 1/6 de la calificación, valorado por el uso correcto de particiones (train/valid/test), *cross-validation*, prevención de *data leakage* y trazabilidad del pipeline.
- **Reproducibilidad:** 1/6 de la calificación, valorado por el uso de proyectos, semillas, dependencias (opcional **renv**), y reporte ejecutable (Quarto/RMarkdown) con instrucciones claras.
- **Calidad del código:** 1/6 de la calificación, valorado por legibilidad, modularidad, uso de funciones, buenas prácticas y estructura de proyecto.

- **Análisis crítico:** 1/6 de la calificación, valorado por la interpretación de resultados, justificación de métricas/modelos, identificación de limitaciones, sesgos y riesgos (p. ej., desbalance, sesgo de selección).
- **Comunicación técnica:** 1/6 de la calificación, valorado por claridad del reporte, figuras/tablas informativas, conclusiones y recomendaciones accionables.

4 Bibliografía

- Duda, R.O., Hart, P.E. & Stork, D.G. (2001) *Pattern Classification* (2nd Edition) Wiley.
- Bishop, C.M. (2006) *Pattern Recognition and Machine Learning*. Springer.
- Hastie, T., Tibshirani, R. & Friedman, J. (2009) *The Elements of Statistical Learning* (2nd Edition). Springer.
- James, G., Witten, D., Hastie, T. & Tibshirani, R. (2021) *An Introduction to Statistical Learning* (2nd Edition). Springer.
- Murphy, K.P. (2012) *Machine Learning: A Probabilistic Perspective*. MIT Press.
- Kuhn, M. & Johnson, K. (2019) *Feature Engineering and Selection*. CRC Press.
- Wickham, H. & Golemund, G. (2017) *R for Data Science* (2nd Ed. disponible en línea). O'Reilly.
- Kuhn, M. & Silge, J. (2022) *Tidy Modeling with R*. O'Reilly.
- Artículos selectos recientes (biomédico, visión, texto, sensores) de revistas y congresos de reconocimiento de patrones.