

Métodos Estadísticos de Predicción

Guía de la
Asignatura

Grado en Matemáticas

AUTORES

- Víctor Aceña Gil
- Isaac Martín de Diego

2025-2026



Índice de Contenidos

Introducción a la asignatura	3
• Objetivos de la asignatura	3
• Temario de la asignatura	4
Desarrollo de la asignatura	5
Bibliografía	7
• Bibliografía Principal y Recomendada	7
• Referencias Adicionales y Tópicos Relacionados	8

Introducción a la asignatura

Los modelos estadísticos han emergido como herramientas fundamentales en la era de la información, donde la capacidad de analizar y predecir comportamientos a partir de datos se ha convertido en una habilidad esencial. En este contexto, los modelos para la predicción juegan un papel crucial al permitirnos describir y cuantificar las relaciones entre variables, así como anticipar resultados futuros. Este libro está diseñado para proporcionar una comprensión profunda y práctica de estas técnicas, basándose en el contenido de la asignatura impartida en el **Grado en Matemáticas**.

El modelado estadístico tiene un propósito dual que exploraremos en profundidad: la **predicción** (la búsqueda de la máxima precisión) y la **inferencia** (la búsqueda de la explicación y la comprensión). Entender esta dicotomía es clave para seleccionar y evaluar los modelos de manera efectiva.

Comenzaremos sentando las bases de la regresión lineal, desde su formulación más simple hasta la complejidad del modelo múltiple. A partir de ahí, aprenderemos a enriquecer nuestros modelos mediante la **ingeniería de características**, a seleccionar las variables más relevantes con técnicas de **selección y regularización**, y finalmente, a extender nuestro alcance más allá de la normalidad con los **Modelos Lineales Generalizados (GLM)**.

Objetivos de la asignatura

- Aprender a formular, estimar e interpretar modelos de regresión, comprendiendo los supuestos teóricos que los sustentan.
- Dominar el diagnóstico de los modelos para verificar su validez y detectar problemas como la multicolinealidad o la heterocedasticidad.
- Saber aplicar técnicas de ingeniería de características, selección de variables y regularización para optimizar el rendimiento y la interpretabilidad de los modelos.
- Extender el conocimiento de los modelos lineales a los GLM para poder modelar diferentes tipos de variables de respuesta (binarias, de conteo, etc.).
- Validar rigurosamente la capacidad predictiva de los modelos para asegurar su generalización a nuevos datos.

Temario de la asignatura

- **Tema 1:** Introducción a los Modelos de Regresión
- **Tema 2:** El Modelo de Regresión Lineal Simple
- **Tema 3:** El Modelo de Regresión Lineal Múltiple
- **Tema 4:** Ingeniería de Características: Transformaciones e Interacciones
- **Tema 5:** Selección de Variables, Regularización y Validación
- **Tema 6:** Modelos de Regresión Generalizada (GLM)
- **Tema 7:** Tópicos Avanzados en Modelado Predictivo

Desarrollo de la asignatura

El curso se desarrolla a lo largo de 15 semanas, con dos sesiones semanales. La metodología busca integrar la teoría y la práctica de forma fluida: los conceptos teóricos se presentarán junto con su implementación en R, fomentando un aprendizaje aplicado en cada clase. Los apuntes de la asignatura incluyen todo el código necesario para que el alumno pueda seguir las explicaciones y desarrollar las prácticas correspondientes.

- **Semana 1: Presentación e Introducción a los Modelos de Regresión (Tema 1)**
 - Presentación de la asignatura, guía docente y sistema de evaluación. Se abordarán los conceptos fundamentales del modelado estadístico: el propósito dual de predecir vs. explicar y la anatomía de un modelo de regresión.
- **Semana 2: Regresión Lineal Simple (RLS) - Fundamentos y Estimación (Tema 2)**
 - Se cubrirá la formulación del modelo, los supuestos teóricos y la estimación de parámetros mediante Mínimos Cuadrados Ordinarios (MCO), junto con la interpretación de coeficientes y R^2 .
- **Semana 3: RLS - Inferencia y Diagnóstico (Tema 2)**
 - Se profundizará en la evaluación del modelo mediante la inferencia sobre los coeficientes (test t), ANOVA (test F) y el diagnóstico completo de los supuestos a través del análisis de residuos.
- **Semana 4: Regresión Lineal Múltiple (RLM) - Formulación e Interpretación (Tema 3)**
 - Extensión al caso multivariante, con especial énfasis en la interpretación de coeficientes bajo el principio ceteris paribus y la introducción al problema de la multicolinealidad.
- **Semana 5: Ingeniería de Características I - Transformaciones (Tema 4)**
 - Se explorarán técnicas para mejorar el modelo, como las transformaciones de variables (logarítmica, Box-Cox) y la codificación de variables categóricas.
- **Semana 6: Ingeniería de Características II - Interacciones (Tema 4)**
 - Se estudiará cómo modelar relaciones más complejas mediante la inclusión, visualización e interpretación de términos de interacción entre variables.
- **Semana 7: Consolidación y Práctica de Regresión Lineal**
 - Semana dedicada a ejercicios prácticos completos para afianzar los conceptos de RLS, RLM y la ingeniería de características, sirviendo como preparación para el Trabajo 1.

- **Semana 8: Evaluación Parcial e Introducción a la Selección de Variables**
 - Defensa oral del Trabajo 1. Se comenzará el siguiente bloque temático con la introducción a los criterios de información (AIC, BIC, Cp de Mallows).
- **Semana 9: Métodos de Selección de Variables (Tema 5)**
 - Estudio de los métodos automáticos (stepwise) y exhaustivos (best subset), analizando en profundidad sus ventajas y, sobre todo, sus limitaciones prácticas.
- **Semana 10: Regularización y Validación de Modelos (Tema 5)**
 - Se introducirán las técnicas de regularización (Ridge, Lasso, Elastic Net) y los fundamentos de la validación de modelos, con especial foco en la validación cruzada (Cross-Validation).
- **Semana 11: Modelos Lineales Generalizados (GLM) - Fundamentos (Tema 6)**
 - Introducción a los componentes clave que extienden el modelo lineal: la familia exponencial, la función de enlace y la estimación por Máxima Verosimilitud (MLE).
- **Semana 12: GLM - Regresión Logística (Tema 6)**
 - Estudio en profundidad del modelo para respuestas binarias, incluyendo la interpretación de Odds Ratios y las métricas de validación específicas como la curva ROC y el AUC.
- **Semana 13: GLM - Regresión de Poisson (Tema 6)**
 - Se abordará el modelado de datos de conteo, la interpretación de Incidence Rate Ratios (IRR), y el diagnóstico y solución del problema de la sobredispersión mediante la Regresión Binomial Negativa.
- **Semana 14: Tópicos Avanzados y Repaso**
 - Breve introducción a extensiones de los modelos vistos, como los Modelos Aditivos Generalizados (GAMs). Repaso general de toda la materia y sesión de consultas para el examen y el trabajo final.
- **Semana 15: Presentación de Proyectos Finales**
 - Exposición y defensa del Trabajo Colectivo 2 por parte de los alumnos, donde aplicarán las técnicas de la asignatura a un problema de modelado real.

Bibliografía

A continuación, se presenta la bibliografía utilizada y recomendada para el curso, dividida en textos principales y referencias complementarias.

Bibliografía Principal y Recomendada

- Agresti, A. (2015). Foundations of linear and generalized linear models. John Wiley & Sons.
- Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. Journal of the Royal Statistical Society: Series B, 26(2), 211-243.
- Draper, N. R. (1998). Applied regression analysis. McGraw-Hill.
- Fox, J., & Weisberg, S. (2018). An R companion to applied regression. Sage publications.
- Harrell, F. E. (2015). Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis. Springer.
- Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction. Springer.
- Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression. John Wiley & Sons.
- Jaccard, J., & Turrisi, R. (2003). Interaction effects in multiple regression. Sage.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An Introduction to Statistical Learning with Applications in R. Springer.
- Kuhn, M., & Johnson, K. (2019). Feature engineering and selection: a practical approach for predictive models. CRC Press.
- Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). Applied linear statistical models. McGraw-Hill.
- McCullagh, P. (2019). Generalized linear models. Routledge.
- Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized linear models. Journal of the Royal Statistical Society Series A, 135(3), 370-384.
- Pinheiro, J. C., & Bates, D. M. (2000). Mixed-Effects Models in S and S-PLUS. Springer.
- Shmueli, G. (2010). To Explain or to Predict?. Statistical Science, 25(3), 289-310.

- Tukey, J. W. (1977). Exploratory data analysis. Reading, MA.
- Weisberg, S. (2005). Applied linear regression. Wiley.
- Wood, S. N. (2017). Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC.
- Zheng, A., & Casari, A. (2018). Feature engineering for machine learning: principles and techniques for data scientists. O'Reilly Media.

Referencias Adicionales y Tópicos Relacionados

- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1-48.
- Bellman, R. E. (1978). Artificial intelligence: can computers think?.
- Flach, P. (2012). Machine learning: the art and science of algorithms that make sense of data. Cambridge University Press.
- Kelleher, J. D., & Tierney, B. (2018). Data science. MIT Press.
- Kelleher, J. D., Mac Namee, B., & D'arcy, A. (2020). Fundamentals of machine learning for predictive data analytics. MIT press.
- Molnar, C. (2020). Interpretable machine learning. Lulu.com.
- Murphy, K. P. (2012). Machine learning: a probabilistic perspective. MIT press.
- Russell, S. J. (2010). Artificial intelligence a modern approach. Pearson Education, Inc.
- Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining, 1, 29-39.