

Tarea 3: Regresión Lineal y Bayesiana

Dr. Marco Antonio Aquino López
Maestría en Probabilidad y Estadística, CIMAT

Agosto–Diciembre 2025
Fecha de entrega: Lunes 13 de Octubre de 2025

Instrucciones Generales

Resuelva cuidadosamente cada apartado. Todas las demostraciones deben presentarse paso a paso, con claridad en las justificaciones matemáticas y con una breve interpretación estadística de los resultados.

1 Regresión lineal ordinaria (OLS)

1. **Derivación del estimador OLS:** Partiendo del modelo clásico

$$y = X\beta + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2 I),$$

demuestre que el estimador de Mínimos Cuadrados Ordinarios es

$$\hat{\beta} = (X^\top X)^{-1} X^\top y,$$

siempre que $X^\top X$ sea invertible.

2. **Propiedades del estimador:** Calcule explícitamente:

$$E[\hat{\beta}], \quad \text{Var}(\hat{\beta}).$$

Concluya que $\hat{\beta}$ es insesgado y eficiente dentro de la clase de estimadores lineales (teorema de Gauss–Markov).

2 Regresión lineal bayesiana (prior conjugado)

1. **Prior conjugado:** Suponga un prior conjugado:

$$\beta \mid \sigma^2 \sim \mathcal{N}(\beta_0, \sigma^2 V_0), \quad \sigma^2 \sim \text{Inv-Gamma}(a_0, b_0).$$

2. **Distribución posterior:** Derive los parámetros posteriores (β_n, V_n, a_n, b_n) y escriba la forma explícita de la posterior conjunta

$$p(\beta, \sigma^2 \mid y).$$

3. **Distribuciones marginales:** Identifique las distribuciones marginales de β y de σ^2 .

3 Conexión con regularización

1. **Regresión Ridge:** Muestre que si se toma un prior Normal isotrópico

$$\beta \sim \mathcal{N}(0, \tau^2 I),$$

el estimador de máxima a posteriori (MAP) es equivalente a la regresión Ridge:

$$\hat{\beta}_{\text{MAP}} = \arg \min_{\beta} \|y - X\beta\|^2 + \lambda \|\beta\|^2, \quad \lambda = \sigma^2 / \tau^2.$$

2. **Regresión Lasso:** Muestre que si en lugar de un prior Normal se utiliza un prior Laplace (doble-exponencial)

$$p(\beta_j) \propto \exp(-\lambda |\beta_j|),$$

el estimador MAP corresponde a la regresión Lasso:

$$\hat{\beta}_{\text{MAP}} = \arg \min_{\beta} \|y - X\beta\|^2 + \lambda \|\beta\|_1.$$

4 Extensiones: errores no normales

1. **Modelos alternativos:** Proponga un modelo de regresión donde el error ε no siga una distribución Normal. Ejemplos:
 - $\varepsilon \sim \text{Laplace}(0, b)$ (robusto a outliers).
 - $\varepsilon \sim \text{Student-}t(\nu)$ (colas pesadas).
2. **Consecuencias metodológicas:** Explique cuáles serían las consecuencias sobre:
 - La forma de la verosimilitud.
 - La existencia o no de priors conjugados.
 - Los métodos de inferencia requeridos (MCMC, aproximación variacional, etc.).

Análisis de Artículos Científicos

Objetivo

Reconocer, criticar y replicar el uso de modelos de regresión en trabajos científicos reales. Cada estudiante deberá seleccionar y analizar dos artículos científicos publicados en revistas indexadas: **uno que utilice regresión lineal y otro que utilice regresión logística.**

5 Selección de artículos

1. Busque y seleccione **dos artículos científicos de los últimos 5 años** en su área de interés donde:
 - **Artículo 1:** Utilice **regresión lineal**
 - **Artículo 2:** Utilice **regresión logística**
2. Justifique por qué eligió esos artículos y en qué contexto aplican los modelos de regresión.
3. Anexe las referencias completas en formato APA.

6 Análisis crítico

1. Describa cómo se utilizaron los modelos de regresión en cada artículo:
 - Variables dependientes e independientes.
 - Supuestos implícitos (linealidad, homocedasticidad, normalidad, independencia).
 - Interpretación de los coeficientes.
2. Discuta si considera que el modelo de regresión fue el método adecuado en cada caso.
3. Señale limitaciones del análisis y proponga alternativas (ej. modelos robustos, transformaciones, regresión bayesiana, etc.).

7 Replicación de resultados

1. Obtenga los datos originales:
 - Descargándolos de un repositorio asociado al artículo (si está disponible), o
 - Capturándolos a partir de gráficas o tablas publicadas.
2. Replique el análisis de regresión reportado en el artículo.
3. Compare sus resultados con los publicados y discuta posibles diferencias.

8 Exploración de alternativas

1. Aplique un análisis alternativo de regresión sobre los mismos datos. Ejemplos:
 - Para regresión lineal: Ridge, Lasso o regresión robusta.
 - Para regresión logística: Regularización L1/L2 o modelos Bayesianos.
2. Valide el modelo ajustado mediante:
 - Gráficos de residuos.

- Medidas de bondad de ajuste (AIC, BIC, R^2 ajustado).
 - Validación cruzada.
3. Compare el desempeño de su propuesta con el modelo original.

Entregable

- Un informe en PDF (máx. 10 páginas) que incluya:
 1. Descripción de los artículos seleccionados.
 2. Análisis crítico del uso de los modelos de regresión.
 3. Replicación del análisis original con resultados comparativos.
 4. Exploración de un modelo alternativo y su validación.
 5. Conclusiones generales.
- Archivos de código en Python (bien comentados y reproducibles).
- Archivos de datos utilizados (si fueron reconstruidos, explique el procedimiento).
- Un repositorio en **GitHub**, gestionado por equipos de 3 estudiantes (equipos previamente asignados al azar), que contenga:
 - El informe en PDF.
 - El código reproducible.
 - Los datos y/o scripts de reconstrucción.
 - Un archivo `README.md` con instrucciones claras para ejecutar el proyecto.