APRENDIZAJE AUTOMÁTICO I

UDESA – Maestría en IA - 2025

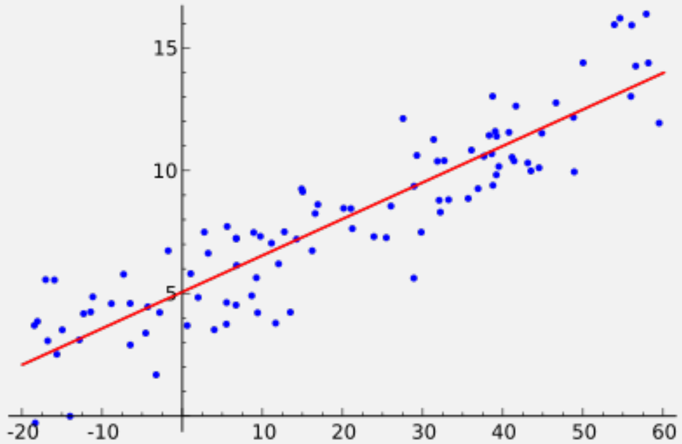
MATIAS LEONI

CLASE 2: REGRESIÓN LINEAL

- **Agenda de Hoy:**
  - Introducción y Conceptos Fundamentales de la Regresión Lineal
  - La Función de Hipótesis y la Función de Costo (MSE)
  - El Método de Mínimos Cuadrados Ordinarios (OLS)
  - Optimización con Gradiente Descendente (GD y SGD)
  - Regularización para evitar el sobreajuste: Ridge (L2) y Lasso (L1)
  - Selección de Hiperparámetros mediante Validación Cruzada
  - Métricas de Desempeño: MSE, MAE y R^2

FUNDAMENTOS Y MÍNIMOS CUADRADOS

- Comprensión de la intuición detrás de la regresión.
- Formalización del modelo y cómo encontrar los "mejores" parámetros.
- Evaluación de la calidad de nuestro modelo.



¿QUÉ ES LA REGRESIÓN LINEAL?

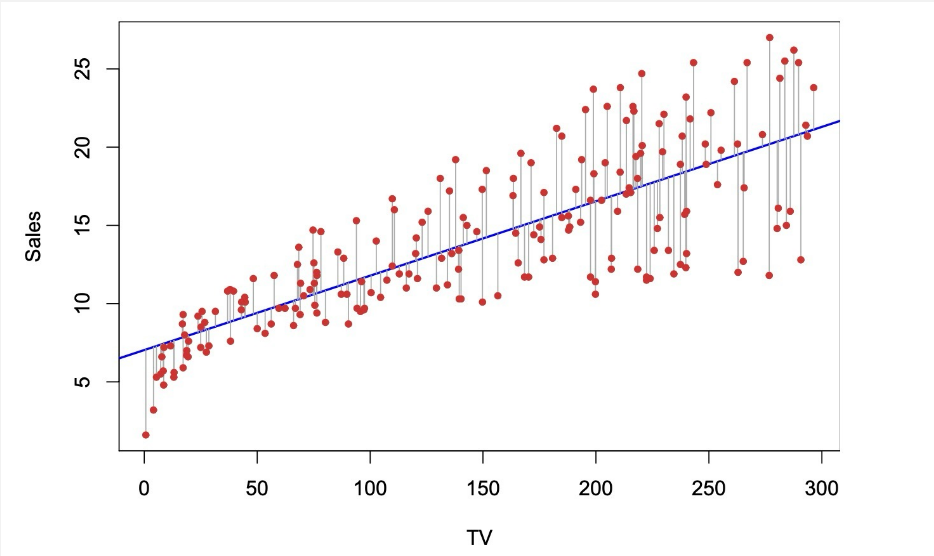
- Es un método de aprendizaje supervisado para predecir una respuesta cuantitativa (continua), como el precio de una casa o las ventas de un producto.
- Asume una relación aproximadamente lineal entre las variables predictoras (X) y la variable de respuesta (Y).
- **Regresión Lineal Simple:** Utiliza un solo predictor. Su ecuación es

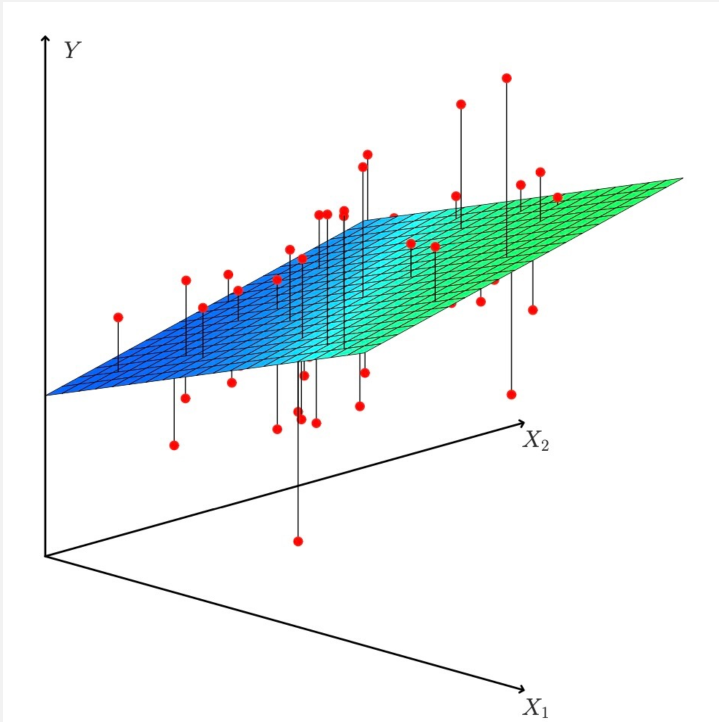
  $$\mathrm{Y} \approx \beta_0 + \beta_1 X$$

- **Regresión Lineal Múltiple:** Utiliza dos o más predictores. Su ecuación es

  $$Y \approx \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$$

¿QUÉ ES LA REGRESIÓN LINEAL?





EL MODELO: HIPÓTESIS Y COEFICIENTES

- La función de hipótesis, $\mathrm{h}(X)$, es el modelo que proponemos para aproximar Y: $\mathrm{h}(X) = \hat{y} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_p$
- Los **coeficientes** o **parámetros** $(\beta_0, \beta_1, \dots, \beta_p)$ son los valores que el modelo "aprende" de los datos.
- **Interpretación de los coeficientes:**
  - β_0: Es el **intercepto**, el valor esperado de Y cuando todas las X son cero.
  - β_j: Es la **pendiente** para X_j. Representa el cambio promedio en Y por un aumento de una unidad en X_j, manteniendo fijos todos los demás predictores.

LA FUNCIÓN DE COSTO: ERROR CUADRÁTICO MEDIO (MSE)

- Para encontrar los mejores parámetros (β_j), necesitamos medir qué tan bien se ajusta el modelo a los datos.
- El **residuo** (ε_i) es la diferencia entre el valor real y el predicho: $\varepsilon_i = y_i - \widehat{y}_i$.
- La **RSS** es la suma de todos los residuos al cuadrado. Nuestro objetivo es minimizar esta cantidad:
  $\mathrm{RSS} = \sum_{i=1}^{n}(y_i - \widehat{y}_i)^2 = \sum_{i=1}^{n}\left(y_i - \beta_0 - \beta_1\, x_{i1} - \cdots - \beta_p\, x_{ip}\right)^2$
- El **Error Cuadrático Medio (MSE)** es el promedio de esta suma y es la función de costo que optimizaremos:
  $MSE = \frac{1}{n} RSS$

SOLUCIÓN ANALÍTICA: MÍNIMOS CUADRADOS ORDINARIOS (OLS)

- OLS es el método más común para estimar los coeficientes minimizando el RSS.
- Para la regresión lineal simple, existe una solución analítica (fórmulas cerradas) que se obtiene mediante cálculo diferencial:
  $\widehat{\beta_1} = \frac{Cov(x,y)}{Var(x)}$ y $\widehat{\beta_0} = \bar{y} - \widehat{\beta_1}\,\bar{x}$
- Para la regresión múltiple, la solución se expresa de forma compacta usando álgebra matricial, pero el concepto de minimizar RSS es el mismo:
  $\hat{\beta} = \left(\mathrm{X}^{\mathrm{T}}\,\mathrm{X}\right)^{-1}\mathrm{X}^{\mathrm{T}}\,\mathrm{y}$
- Una propiedad clave es que la línea de regresión de mínimos cuadrados siempre pasa por el punto de las medias $(\bar{x}, \bar{y})$.
- Otra propiedad: ¡son estimadores insesgados!

EVALUACIÓN DEL MODELO: COEFICIENTE DE DETERMINACIÓN R^2

- El **Coeficiente de Determinación (R^2)** mide la proporción de la varianza total de la respuesta (Y) que es explicada por el modelo.
- Su fórmula es:

  $R^2 = 1 - \frac{RSS}{TSS}$

  donde TSS (Suma Total de Cuadrados) mide la variabilidad inherente en la respuesta:

  $TSS = \sum_{i=1}^{n}(y_i - \bar{y})^2$

- R^2 siempre tiene un valor entre 0 y 1. Un valor más cercano a 1 indica un mejor ajuste del modelo: "bondad del ajuste"
- En el contexto de la regresión lineal simple, R^2 es idéntico al cuadrado del coeficiente de correlación (ρ^2).

EVALUACIÓN DEL MODELO: R^2_{adj}, MSE Y MAE

- Inflación de R^2 , defino R^2 ajustado: $R^2_{adj} = 1 - (1 - R^2)\frac{n-1}{n-p-1}$
- Mean Squared Error (MSE): $MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2$
  - Mide el error cuadrático promedio. Es muy sensible a los outliers (errores grandes son penalizados fuertemente).
- Mean Absolute Error (MAE): $\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \widehat{y}_i|$
  - Mide el error absoluto promedio. Es más robusto a outliers que el MSE y sus unidades son más fáciles de interpretar.

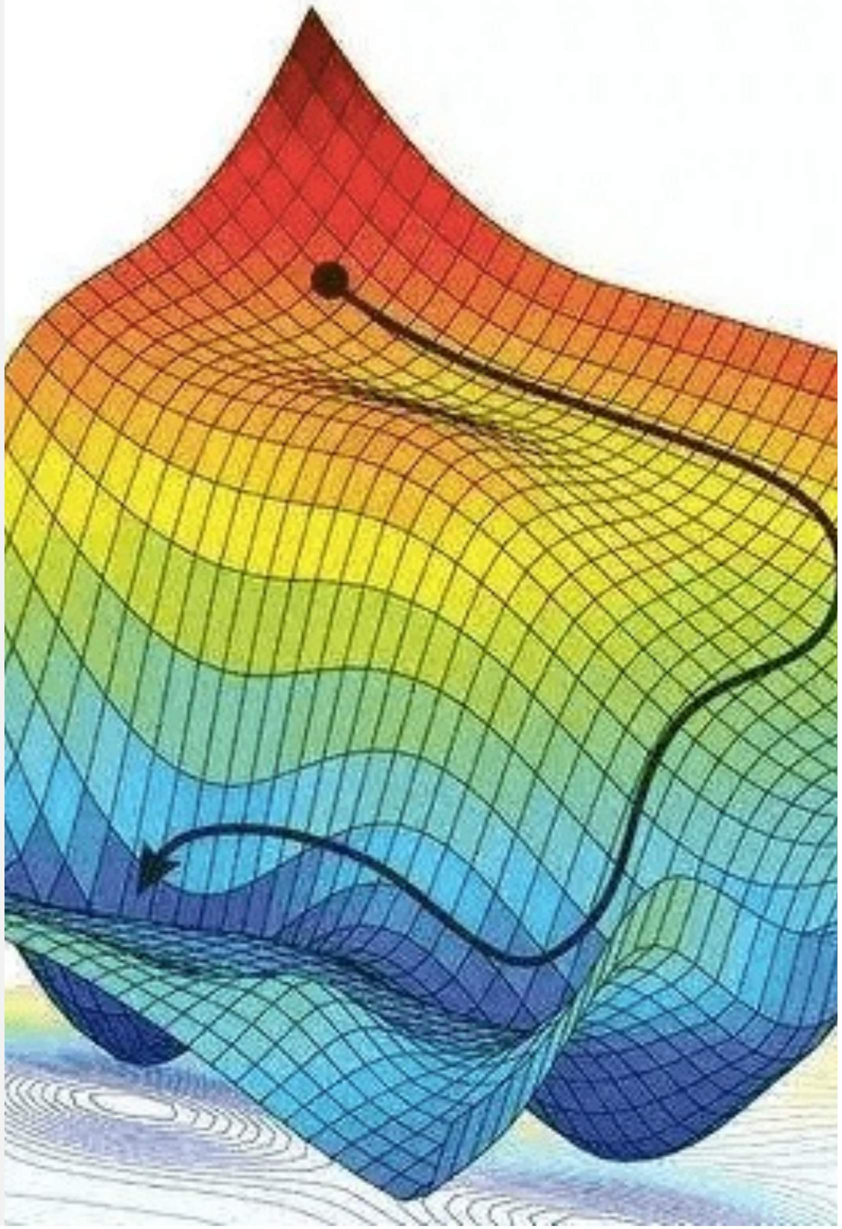
PAUSA (15 MINUTOS)

OPTIMIZACIÓN Y REGULARIZACIÓN

- Un método alternativo para encontrar los coeficientes: Gradiente Descendente.
- Cómo prevenir el sobreajuste en modelos complejos.
- El balance entre sesgo y varianza en la práctica.

SOLUCIÓN NUMÉRICA: GRADIENTE DESCENDENTE

- OLS puede ser computacionalmente costoso si tenemos una cantidad masiva de características.
- El **Gradiente Descendente** es un algoritmo de optimización iterativo para encontrar el mínimo de una función (nuestra función de costo MSE).
- **Intuición:** Imagina que estás en una montaña con los ojos vendados y quieres llegar al punto más bajo. Das un pequeño paso en la dirección donde la pendiente desciende más abruptamente. Repites este proceso hasta llegar al valle.
- **Regla de Actualización:** Ajustamos los coeficientes βj en cada iteración:
  $$\beta_j := \beta_j - \alpha \frac{\partial}{\partial \beta_j} J(\beta_0, \beta_1, \dots)$$
  donde α es la **tasa de aprendizaje (learning rate)**.
- **Descenso de Gradiente Estocástico (SGD):** Una variante que calcula el gradiente usando una sola muestra a la vez, haciéndolo más rápido para datasets muy grandes.



SOBREAJUSTE Y LA NECESIDAD DE REGULARIZACIÓN

- Modelos con muchos predictores pueden volverse muy complejos y "memorizar" el ruido de los datos de entrenamiento en lugar de capturar la tendencia general. Esto es **sobreajuste (overfitting)**.
- Un modelo sobreajustado tiene un rendimiento excelente en los datos de entrenamiento pero muy pobre en datos nuevos (test).
- El sobreajuste a menudo se manifiesta con **coeficientes (β_j) de gran magnitud**.
- La **regularización** es una técnica que busca simplificar el modelo penalizando la magnitud de sus coeficientes para mejorar su capacidad de generalización.

REGULARIZACIÓN: RIDGE (L2) VS. LASSO (L1)

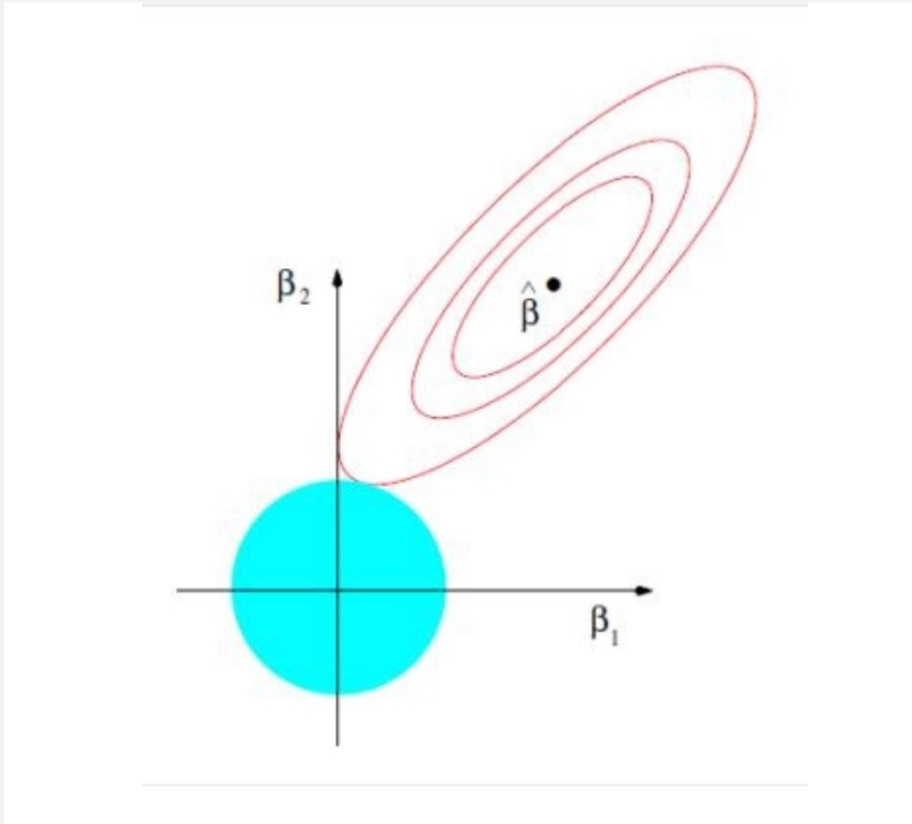
- **Ridge Regression (L2):** Minimiza la función de costo

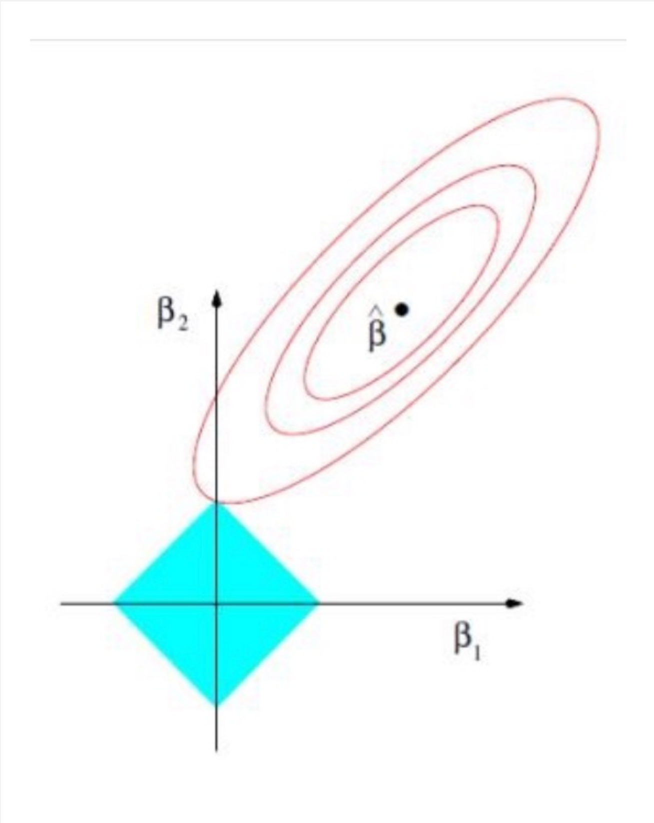
  $RSS + \lambda \sum_{j=1}^{p} \beta_p^2$.

  - La penalización L2 encoge los coeficientes, pero **nunca los hace exactamente cero**.
  - Es útil cuando la mayoría de los predictores son relevantes.
- Solución exacta:

  $$\widehat{\beta} = (X^T X + \lambda I)^{-1} X^T y$$

  - Me salva del problema de la colinealidad.





REGULARIZACIÓN: RIDGE (L2) VS. LASSO (L1)

- **Lasso Regression (L1):** Minimiza la función de costo

  $\mathrm{RSS} + \lambda \sum_{j=1}^{p} |\beta_p|$.

  - La penalización L1 puede forzar a que algunos coeficientes **sean exactamente cero**.
  - Realiza una **selección de características** automática, resultando en un modelo más simple e interpretable.

VALIDACIÓN CRUZADA PARA SELECCIONAR EL MODELO

- La regularización introduce un **hiperparámetro**, λ, que controla la fuerza de la penalización. ¿Cómo elegimos el valor óptimo de λ?
- **No podemos usar el conjunto de test** para ajustar λ, ya que contaminaríamos nuestra evaluación final.
- La **Validación Cruzada (Cross-Validation)** es la técnica estándar para esta tarea.
  - Dividimos el conjunto de entrenamiento en K "pliegues" (folds).
  - Entrenamos el modelo K veces, cada vez usando un pliegue diferente como validación y el resto para entrenar.
  - Calculamos el MSE promedio en los pliegues de validación para un valor de λ dado.
- Repetimos este proceso para varios valores de λ y elegimos aquel que nos dé el menor error de validación promedio.

RESUMEN Y PRÓXIMOS PASOS

- La **Regresión Lineal** modela una respuesta continua a partir de predictores.
- Minimizamos el **MSE** usando **Mínimos Cuadrados** o **Gradiente Descendente**.
- Evaluamos el modelo con métricas como **R^2**, **MSE** y **MAE**.
- La **Regularización (Ridge y Lasso)** combate el sobreajuste y puede realizar selección de características.
- Usamos **Validación Cruzada** para ajustar hiperparámetros como λ.
- **Próximos Pasos:** ¡A laburar! Aplicaremos estos conceptos en un taller práctico con Python en un Jupyter Notebook.