Apuntes semana 5 Clase #1s

Eder Vega Suazo
Escuela de Ingeniería en Computación
Instituto Tecnológico de Costa Rica
IC-6200 - Inteligencia Artificial Gr2

***Resumen*—Este documento es un resumen de la clase de inteligencia artificial correspondiente a la semana 5, enfocando en los fundamentos del aprendizaje supervisado. Se abordan temas clave como la optimización de modelos mediante cálculo diferencial y el algoritmo de descenso de gradiente aplicado a la función de error cuadrático medio. Además, se examinan desafíos comunes en el modelado predictivo, incluyendo el manejo de relaciones no lineales entre variables y la detección de valores atípicos. Tambien discuten estrategias para la evaluación de modelos mediante partición de datasets y se analiza el compromiso entre sesgo y varianza, crucial para desarrollar modelos con capacidad de generalización efectiva.**

I. Optimización mediante cálculo diferencial

I-A. *Función de error cuadrático medio*

En problemas de regresión, la función de costo más común está dada por:

$$L = \frac{1}{N}\sum_{i=1}^{N}\left(f_{w,b}(x_i) - y_i\right)^2, \quad i = 1, \ldots, N$$

donde $h_\theta(x_i)$ representa la predicción del modelo para la instancia i-ésima.

El proceso de optimización busca minimizar esta función mediante el cálculo de gradientes:

$$\frac{\partial L}{\partial w} = \frac{1}{N}\sum_{i=1}^{N} 2((wx_i + b) - y_i) \cdot x_i$$

$$\frac{\partial L}{\partial b} = \frac{1}{N}\sum_{i=1}^{N} 2((wx_i + b) - y_i)$$

I-B. *Algoritmo de descenso de gradiente*

La actualización de parámetros se realiza de forma iterativa mediante:

$$w^{(t+1)} = w^{(t)} - \alpha\frac{\partial L}{\partial w^{(t)}}$$

$$b^{(t+1)} = b^{(t)} - \alpha\frac{\partial L}{\partial b^{(t)}}$$

donde α representa la tasa de aprendizaje que controla la magnitud de cada actualización.

I-C. *Terminología fundamental*

- **Época (Epoch)**: Ciclo completo de presentación de todos los ejemplos de entrenamiento al modelo.
- **Lote (Batch)**: Subconjunto de ejemplos utilizados para calcular una actualización de parámetros.
- **Tasa de aprendizaje**: Hyperparámetro que determina la velocidad de convergencia del algoritmo.

II. Desafíos en modelado predictivo

II-A. *Relaciones no lineales entre variables*

La regresión lineal presume una relación lineal entre predictores y variable respuesta. Cuando esta suposición se viola, el modelo resulta inadecuado y muestra patrones sistemáticos en los residuos:

$$e_i = y_i - \hat{y}_i$$

La solución implica transformar las variables predictoras mediante expansión polinomial o otras transformaciones que permitan capturar relaciones no lineales manteniendo la linealidad en los parámetros.

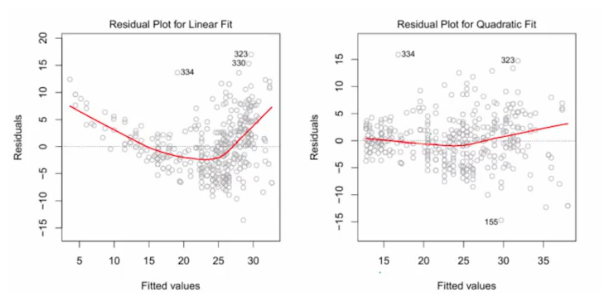

Figura 1: Ejemplo de relación no lineal y su ajuste mediante transformación polinomial.

II-B. *Manejo de valores atípicos*

Las observaciones extremas pueden distorsionar significativamente los modelos de regresión. Existen múltiples enfoques para su identificación y tratamiento:

II-B1. *Identificación de valores atípicos:*

- **Residuos estandarizados**: $z_i = \frac{e_i}{\sigma_e}$ donde σ_e es la desviación estándar de los residuos.
- **Rango intercuartílico**: Valores fuera de $[Q_1 - 1{,}5 \cdot IQR, Q_3 + 1{,}5 \cdot IQR]$ se consideran atípicos.

II-B2. *Técnicas de tratamiento:*

- **Eliminación**: Remover observaciones identificadas como atípicas.
- **Winsorización**: Reemplazar valores extremos por percentiles específicos (ej. percentil 5 y 95).
- **Transformaciones**: Aplicar funciones como logaritmo o raíz cuadrada para reducir la influencia de valores extremos.

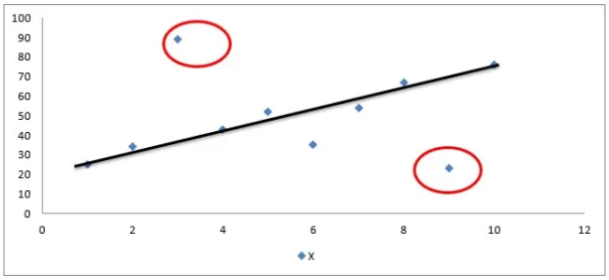

Figura 2: Efecto de valores atípicos en un modelo de regresión lineal.

III. Evaluación y validación de modelos

III-A. Partición de datasets

La división adecuada de los datos es crucial para evaluar la capacidad de generalización:

Cuadro I: Propósitos de los diferentes subconjuntos de datos

Subconjunto	**Propósito**
Entrenamiento	Ajuste de parámetros del modelo mediante optimización
Validación	Selección de hyperparámetros y monitorización del sobreajuste
Prueba	Evaluación final del rendimiento con datos nunca vistos

III-B. Técnicas de muestreo

1. **Muestreo aleatorio**: División randomizada que preserva la distribución original de los datos.
2. **Muestreo estratificado**: Mantiene la proporción de clases en cada partición, crucial para datos desbalanceados.
3. **Validación cruzada**: Divide los datos en k particiones y realiza k iteraciones de entrenamiento/validación.

Figura 3: Esquema de validación cruzada con $k = 5$ particiones.

Cuadro II: Características de modelos con sesgo o varianza elevados

Métrica	**Alto sesgo**	**Alta varianza**
Error entrenamiento	Alto	Bajo
Error validación	Alto	Alto
Comportamiento	Subajuste	Sobreajuste
Soluciones	Modelos más complejos	Regularización, más datos

IV. Sesgo y varianza

IV-A. Diagnóstico de problemas comunes

IV-B. Estrategias de mejora

- **Para alto sesgo**: Aumentar la complejidad del modelo, agregar características adicionales o reducir regularización.
- **Para alta varianza**: Aumentar datos de entrenamiento, aplicar técnicas de regularización o reducir la complejidad del modelo.
- **Compromiso óptimo**: Seleccionar la complejidad del modelo que minimice el error de generalización.

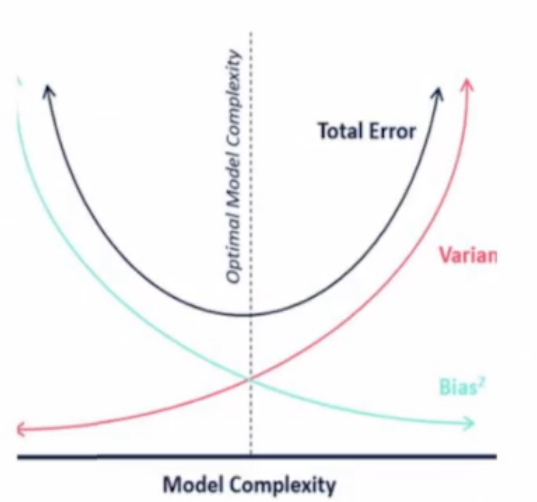

Figura 4: Relación entre complejidad del modelo y error de generalización.

V. Conclusiones

La efectividad de los modelos de aprendizaje supervisado depende críticamente de la adecuada optimización de parámetros, el manejo de relaciones complejas entre variables, la identificación y tratamiento de valores atípicos, y la evaluación rigurosa mediante técnicas de validación apropiadas. El entendimiento del compromiso entre sesgo y varianza permite desarrollar modelos que generalizan efectivamente a nuevos datos, balanceando complejidad y capacidad predictiva.

Referencias

[1] Apuntes de la clase de Inteligencia Artificial, Profesor S. Pacheco, Instituto Tecnológico de Costa Rica, 2025.