

Práctica Calificada 1

Machine Learning para CCSS y Gestión Pública

28 de Octubre de 2024

Instrucciones Generales

- **Grupos:** El trabajo será presentado por grupos de 4 integrantes.
- **Entregable:** Un único Jupyter Notebook (.ipynb), que contenga todo el código, los resultados (gráficos y salidas de celdas) y las respuestas escritas a las preguntas de interpretación.
- **Fecha límite:** La fecha límite para la entrega es el domingo 02 de noviembre las 11:59 pm. Se debe subir su solución a la actividad correspondiente creada en la plataforma de paideia.
- **Cualquier duda:** Por favor, no duden en escribirme por correo a zdelacruzl@pucp.edu.pe.

1. Fundamentos Teóricos y Conceptuales

1.1. Pregunta 1: Sesgo-Varianza y la Regularización de Modelos

Un equipo de politólogos está modelando un fenómeno social complejo, como la participación electoral, utilizando un modelo de regresión flexible que incluye un gran número de variables predictoras. Para evitar el sobreajuste y controlar la complejidad del modelo, deciden utilizar una técnica de regularización (o penalización).

El equipo prueba tres versiones del mismo modelo, que se diferencian únicamente por la intensidad del **parámetro de penalización** aplicado:

- **Versión A (Alta Penalización / Baja Complejidad):** Se aplica una penalización muy fuerte. Esto obliga al modelo a ser muy simple, forzando a que la mayoría de los coeficientes de las variables sean pequeños o exactamente cero. El modelo resultante es muy restringido y se basa en unos pocos predictores clave.
 - **Versión B (Penalización Equilibrada / Complejidad Media):** Se aplica una penalización moderada, seleccionada a través de validación cruzada.
 - **Versión C (Baja Penalización / Alta Complejidad):** La penalización es muy baja o nula. El modelo tiene libertad para usar todas las variables disponibles y ajustarse a los datos de entrenamiento.
- a) **Análisis del Sesgo:** El **sesgo** se define como el error que se introduce al aproximar un problema de la vida real, que puede ser muy complicado, por un modelo mucho más simple. Formalmente, es la diferencia entre la predicción promedio de nuestro modelo y la función verdadera que intentamos estimar ($E[\hat{f}(x)] - f(x)$). Ordene las tres versiones del modelo (de menor a mayor) según su **sesgo** esperado y justifique su orden.
- b) **Análisis de la Varianza:** La **varianza** se refiere a la cantidad en que la estimación de nuestro modelo (\hat{f}) cambiaría si lo entrenáramos con un conjunto de datos de entrenamiento diferente. Mide la sensibilidad del modelo a las particularidades de la muestra. Ordene las tres versiones del modelo (de menor a mayor) según su **varianza** esperada y justifique su orden.
- c) **El Dilema del Investigador:** Un colega argumenta que la Versión C es la mejor porque tiene el error más bajo en los datos de entrenamiento. Explique por qué este enfoque es erróneo para evaluar el rendimiento real del modelo. Dibuje un gráfico conceptual que ilustre su punto, mostrando las curvas del error de entrenamiento (Train Error) y el error de prueba (Test Error) a medida que la complejidad

del modelo aumenta (es decir, a medida que la penalización disminuye). Ubique aproximadamente las Versiones A, B y C en el eje de complejidad.

1.2. Pregunta 2: Estrategias de Validación en Investigación Social

Una socióloga está desarrollando un modelo para predecir el nivel de “confianza en las instituciones” (una escala de 0 a 100) en ciudadanos de un país. Dispone de una encuesta nacional de alta calidad pero con un número limitado de encuestados: **N = 600** observaciones.

Está considerando dos métodos de validación cruzada para estimar el error de su modelo:

- **Método A:** Validación Cruzada de 5 pliegues (5-fold CV).
- **Método B:** Validación Cruzada Dejando Uno Fuera (LOOCV).

Usted debe asesorarla en su decisión.

- a) **Análisis de la Calidad de la Estimación del Error:** Para que su artículo sea creíble, la socióloga quiere una estimación del error que sea lo más cercana posible al “verdadero” error de generalización (es decir, una estimación con **bajo sesgo**). ¿Qué método (A o B) produce una estimación menos sesgada y por qué?
- b) **Recomendación Metodológica:** ¿Qué método le recomendaría a la socióloga para su publicación? Justifique su elección, argumentando por qué es la opción más equilibrada y defendible.

2. Ejercicio Práctico: Descenso de Gradiente

En este ejercicio, implementaremos el **Descenso de Gradiente (En este caso del tipo Batch Gradient Descent)** para encontrar el mínimo de una función de dos variables. La función a optimizar es:

$$f(x, y) = x^2 + 2y^2 - 2xy + 2x - 8y + 10$$

El mínimo teórico se encuentra en $(x = 2, y = 3)$.

2.1. Definición de la Función y su Gradiente

Si necesita ayuda con la estructura del código, consulte el documento de guía adjunto.

- a) Cree una función en Python `f(punto)` que devuelva el valor de la función $f(x, y)$.
- b) Cree una función `grad_f(punto)` que devuelva el gradiente $\nabla f(x, y) = \begin{pmatrix} 2x - 2y + 2 \\ 4y - 2x - 8 \end{pmatrix}$.

2.2. Implementación del Algoritmo

- a) Cree una función `descenso_gradiente(punto_inicial, alpha, n_iteraciones)` que implemente el bucle de optimización y devuelva la trayectoria de los puntos.

2.3. Visualización y Verificación (Caso Base)

- a) Defina: `punto_inicial = np.array([-5.0, 10.0])`, `alpha = 0.1`, `n_iteraciones = 50`.
- b) Ejecute su función y genere un gráfico de contorno de $f(x, y)$.
- c) Sobre el gráfico, dibuje la trayectoria del algoritmo (se recomienda usar flechas).
- d) **Pregunta:** ¿El algoritmo converge visualmente hacia el mínimo teórico?

2.4. Análisis Comparativo del Algoritmo

- a) **Efecto de Tasa de Aprendizaje Alta:** Repita el punto anterior con `alpha = 0.3`.
 - **Pregunta:** ¿Qué comportamiento observa en la trayectoria? ¿Por qué ocurre?
Hint: Recuerde la regla de actualización del Descenso de Gradiente:

$$\mathbf{p}_{\text{nuevo}} = \mathbf{p}_{\text{viejo}} - \alpha \cdot \nabla f(\mathbf{p}_{\text{viejo}})$$

La tasa de aprendizaje α actúa como un “multiplicador” del paso. ¿Qué sucede si este multiplicador es demasiado grande? Piense en la analogía de bajar una colina dando saltos muy largos.

b) **Efecto de Tasa de Aprendizaje Baja:** Repita con $\alpha = 0.01$.

- **Pregunta:** Compare la distancia recorrida con el caso base. ¿Cuál es la desventaja de usar una tasa de aprendizaje tan baja?

3. Ejercicio Práctico: Regresión Lasso y Regularización

Usaremos el dataset **Boston Housing** para predecir el valor mediano de las viviendas (`medv`) utilizando Regresión Lasso.

URL: <https://raw.githubusercontent.com/qlabpucp/datasets/main/datasets/boston.csv>

3.1. Carga y Preparación de Datos

- a) Cargue el dataset desde la URL en un DataFrame.
- b) Explora la base de datos y realice una limpieza si en caso lo amerite.
- c) Separe los datos en predictores (**X**) y objetivo (**y**, la columna `medv`).
- d) Estandarice las características en **X** usando `StandardScaler`.

3.2. Implementación de Validación Cruzada con Lasso

- a) Configure y entrene un objeto `LassoCV` con `cv=10` y `random_state=42`.
- b) Imprima el valor óptimo de α encontrado.

3.3. Interpretación del Gráfico Sesgo-Varianza

- a) Genere el gráfico que muestra el MSE promedio vs. los valores de α (en escala logarítmica).
- b) **Región de Overfitting (izquierda):** Explique por qué un α muy pequeño resulta en un modelo con alta varianza.
- c) **Región de Underfitting (derecha):** Explique por qué un α muy grande resulta en un modelo con alto sesgo.