402 MÉTODO DE DESCENSO DE GRADIENTE - TAREA 1

September 29, 2025

Pregunta 1

La función de Himmelblau es un estándar para probar algoritmos de optimización, ya que posee cuatro mínimos locales idénticos. El objetivo es implementar y comparar el rendimiento del **método de descenso de gradiente (GD)** estándar con el **descenso de gradiente con momento (GDM)**. Se debe analizar cómo la elección de los hiperparámetros (tasa de aprendizaje y momento) afecta la convergencia.

La función está definida como:

$$f(x, y) = (x^2 + y - 11)^2 + (x + y^2 - 7)^2$$

Actividades:

1. **Implementación:** Defina la función de Himmelblau y su gradiente. Implemente el algoritmo de GD y GDM.
2. **Experimentación:**
   - Ejecute ambos algoritmos desde el punto inicial $(-2.5, 2.5)$.
   - Para GD, pruebe con tasas de aprendizaje $\alpha = [0.001, 0.005, 0.01]$.
   - Para GDM, utilice una tasa de aprendizaje fija $\alpha = 0.005$ y pruebe con coeficientes de momentos $\gamma = [0.5, 0.7, 0.9]$.
   - Utilice un máximo de `10000` iteraciones y una tolerancia de `1e-6` como criterios de detención.
3. **Análisis y visualización:**
   - Genere una tabla comparativa que muestre para cada configuración: el algoritmo, los hiperparámetros usados, el mínimo encontrado, el valor de la función en ese mínimo y el número de iteraciones requeridas.
   - Genere un gráfico de contorno de la función de Himmelblau. Sobre este, grafique las trayectorias de convergencia para la mejor configuración de GD y la mejor configuración de GDM.
   - Escriba una conclusión breve explicando qué algoritmo convergió más eficientemente y por qué el momento ayuda (o perjudica) en este caso particular.

Pregunta 2

Una tasa de aprendizaje fija no siempre es óptima. Las estrategias de decaimiento pueden acelerar la convergencia y evitar oscilaciones cerca del mínimo. El objetivo, es implementar y comparar tres estrategias diferentes de decaimiento de la tasa de aprendizaje para minizar la **función de Rosenbrock**:

$$f(x, y) = (1 - x)^2 + 100(y - x^2)^2$$

Actividades:

1. **Implementación:**
    - Modifique el algoritmo de descenso de gradiente para incorporar un planificador de tasa de aprendizaje.
    - Implemente las siguientes tres estrategias de decaimiento para α_k en la iteración k:
        1. **Decaimiento por tiempo:** $\alpha_k = \frac{\alpha_0}{1+k\cdot\tau}$.
        2. **Decaimiento por pasos:** $\alpha_k = \alpha_0 \cdot d^{\lfloor k/s \rfloor}$, donde d es un factor de decaimiento y s es el tamaño del paso.
        3. **Decaimiento exponencial:** $\alpha_k = \alpha_0 \cdot e^{-k\cdot\tau}$
2. **Experimentación:**
    - Minimice la función de Rosenbrock partiendo desde $(-1, -1)$.
    - Utilice los siguientes parámetros: $\alpha_0 = 0.002$, `max_iter = 10000`.
    - Para el decaimiento por tiempo y exponencial, use: $\tau = \frac{\alpha_0}{10000}$.
    - Para el decaimiento por pasos, use un factor $d = 0.5$ y un tamaño de paso $s = 2000$.
    - Compare estos tres métodos con el descenso de gradiente con α fijo de 0.002
3. **Análisis y visualización:**
    - Genere una tabla comparando las cuatro estrategias (incluida la de α fijo), mostrando el punto final alcanzado y el número de iteraciones necesarias para converger.
    - Cree un gráfico que muestre la evolución del valor de la función objetivo $f(x, y)$ a lo largo de las iteraciones para cada una de las cuatro estrategias. Use una escala logarítmica en el eje Y para una mejor visualización.
    - En un segundo gráfico, muestre cómo cambia el valor de la tasa de aprendizaje α_k en cada iteración para las tres estrategias de decaimiento.
    - Concluya cuál estrategia de decaimiento funcionó mejor para este problema y por qué.

Pregunta 3

El método de mínimos cuadrados usado en la regresión lineal es muy sensible a los valores atípicos (outliers), ya que el error se eleva al cuadrado, magnificando el impacto de los puntos lejanos. Una alternativa más robusta es minimizar el **Error Absoluto Medio (MAE)**.

La función objetivo MAE es:

$$f(\beta_0, \beta_1) = \frac{1}{m} \sum_{i=1}^{m} |y_i - (\beta_0 + \beta_1 x_i)|$$

Esta función no es diferenciable en todos sus puntos debido a la función de valor absoluto. Sin embargo, podemos usar su **subgradiente** para aplicar el descenso de gradiente.

El subgradiente de la función objetivo es:

$$\frac{\partial f}{\partial \beta_j} = \frac{1}{m} \sum_{i=1}^{m} x_{ij} \cdot \mathrm{sign}(y_i - \hat{y}_i)$$

donde $\mathrm{sign}(z)$ es -1 si $z < 0$, $+1$ si $z > 0$, y 0 si $z = 0$.

El objetivo es implementar un algoritmo de regresión lineal robusto utilizando el descenso de subgradiente para minimizar el MAE y compararlo con el modelo estándar de Mínimos Cuadrados en presencia de datos atípicos.

Actividades:

1. **Datos:**
   - Utilice los datos de `frecuencia_reloj` y `consumo_energia` del apunte de clases.
   - Introduzca dos valores atípicos (outliers) en los datos para simular errores de medición, para la frecuencia de reloj $[2.5, 4.5]$ y para el consumo de energia $[130, 70]$.
2. **Implementación:**
   - Defina la función objetivo MAE y su subgradiente.
   - Implemente un algoritmo de descenso de gradiente que utilice el subgradiente MAE para encontrar los parámetros β óptimos.
3. **Experimentación y análisis:**
   - Ejecute el algoritmo de regresión lineal original (basado en MSE) sobre el conjunto de datos **con outliers**.
   - Ejecute el nuevo algoritmo de regresión robusta (basado en MAE) sobre el mismo conjunto de datos **con outliers**.
   - **Gráfico:** Cree un único diagrama de dispersión que muestre los datos con los outliers. Sobre este gráfico, dibuje las dos rectas de regresión obtenidas:
     1. La recta del modelo de Mínimos Cuadrados (MSE).
     2. La recta del modelo de Error Absoluto Medio (MAE).
   - **Análisis:** En una breve conclusión, explique cuál de las dos rectas se ajusta mejor a la tendencia general de los datos originales. ¿Por qué el modelo basado en MAE es considerado más “robusto” ante valores atípicos? ¿Cómo se relaciona esto con la penalización que cada función de objetivo impone a los errores grandes?

Juan F. Olivares Pacheco (jfolivar@uda.l)

Universidad de Atacama, Facultad de Ingeniería, Departamento de Matemática