Métodos numéricos y Optimización - primer semestre 2024

Trabajo Práctico 4 - Optimizacion

Fecha de entrega: 30 noviembre 23h59

Escribir un informe reportando los resultados de los siguientes experimentos numéricos. El informe debe contar con una introducción, descripción de los métodos numéricos, análisis de los resultados y conclusiones. El informe puede tener hasta 18 páginas sin contar las referencias y deberá venir junto a los códigos .py (referenciados en el informe).

1. Optimización en 2 dimensiones: gradiente descendiente

En esta sección, trabajaremos con la *función de Rosenbrock* en dos dimensiones, definida como:

$$f(x, y) = (a - x)^2 + b(y - x^2)^2,$$

empleando los valores $a = 1$ y $b = 100$. Esta función es una prueba común en problemas de optimización debido a su forma no convexa, que presenta un valle estrecho que conduce al mínimo global en $(x, y) = (a, a^2)$.

El objetivo es emplear el método de optimización de **gradiente descendente** evaluando su desempeño al encontrar el mínimo de la función.

Gradiente Descendente

- Implementar el algoritmo de gradiente descendente para minimizar $f(x, y)$.
- Probar diferentes tasas de aprendizaje (η o *learning rates*) para observar su impacto en la convergencia:

  $$\mathbf{x}_{n+1} = \mathbf{x}_n - \eta \nabla f(\mathbf{x}_n),$$

  donde $\nabla f(x, y)$ es el gradiente de la función de Rosenbrock, y $\mathbf{x} = (x, y)^t$.

Análisis:

- Estudiar cómo afecta la elección de la tasa de aprendizaje en el gradiente descendente.
- Analizar la sensibilidad de los métodos a las condiciones iniciales, probando múltiples valores iniciales (x_0, y_0), evaluando y visualizando algunas trayectorias en el plano (x, y) representativas.
- Estudien la rapidez con la que se alcanza el mínimo global, y observen el número de iteraciones requeridas para alcanzar una tolerancia fija $\|\nabla f(x, y)\| < \varepsilon$.

Opcional: Método de Newton

- Implementen el método de Newton, que utiliza el gradiente y la matriz Hessiana de $f(x, y)$:

$$\mathbf{x}_{n+1} = \mathbf{x}_n - H^{-1}(\mathbf{x}_n)\nabla f(\mathbf{x}_n),$$

  donde $H(x, y)$ es la matriz Hessiana de $f(x, y)$.

- Estudien cuál método converge más rápido al mínimo global, con qué orden de convergencia.

2. Cuadrados Minimos mediante descenso por gradiente

En esta sección, aplicaremos **gradiente descendente** para resolver un problema de regresión lineal en el dataset *California Housing*, proporcionado por la biblioteca `sklearn`. Este dataset contiene información sobre características demográficas y económicas de diferentes regiones de California, con el objetivo de predecir el valor medio de las viviendas (`MedHouseVal`).

El objetivo principal será encontrar un modelo lineal que mapee las caracteristicas demograficas para inferir el valor medio de las viviendas utilizando los siguientes metodos numericos:

- Cuadrados minimos con **pseudoinversa**, que proporciona una solución analítica exacta para los mínimos cuadrados ordinarios.
- Cuadrados minimos mediante el **gradiente descendente**, que calcula una solución aproximada mediante optimización iterativa.

Descripción del problema

La regresión lineal busca minimizar la funcion objetivo del error cuadrático medio (ECM) de las predicciones $\hat{y} = \mathbf{X}\mathbf{w}$, definido como:

$$\text{ECM}(\mathbf{w}) = \frac{1}{n}\|\mathbf{y} - \mathbf{X}\mathbf{w}\|_2^2,$$

donde:

- $\mathbf{X} \in \mathbb{R}^{n\times(d+1)}$ es la matriz de características con n muestras y d atributos, con una columna extra de 1s para dar cuenta de la ordenada al origen.
- $\mathbf{y} \in \mathbb{R}^n$ es el vector de valores objetivo (valor medio de las viviendas).
- $\mathbf{w} \in \mathbb{R}^{d+1}$ es el vector de coeficientes incognita (variables de decision) a encontrar mediante la optimizacion.

Se debe particionar el dataset en un 80 % de las muestras aleatoriamente como conjunto de entrenamiento y un 20 % como conjunto de testeo. Adicionalmente, las columnas de atributos se deben estandarizar (es decir restarles su media y dividir por la desviación estándar, $\mathbf{v} \rightarrow \frac{\mathbf{v}-\mu}{\sigma}$) calculando las medias y las desviaciones (μ y σ) únicamente a partir de las muestras del conjunto de entrenamiento.

1. **Pseudoinversa**:

- Implementen la solución analítica de mínimos cuadrados utilizando la fórmula:

$$\mathbf{w} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}.$$

(puede escribirla en función de la descomposición SVD de la matriz $\mathbf{X}$).

2. **Gradiente descendente**:

- Implementen el algoritmo de gradiente descendente para minimizar el error cuadrático medio:

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \eta \nabla \text{ECM}(\mathbf{w}_t),$$

¿Es el ECM una función convexa?

- Empleen la tasa de aprendizaje $\eta = 1/\sigma_1^2$, donde σ_1 es el primer valor singular de la matriz $\mathbf{X}$. ¿Por qué tiene sentido utilizar este valor?

Análisis

- Comparen la solución obtenida por la pseudoinversa con la solución iterativa del gradiente descendente para distintos valores de η
- Muestren el error en el conjunto de entrenamiento y de prueba frente al número de iteraciones para gradiente descendente.

Opcional: Regularización L_2

Explorar el impacto de la **regularización** L_2 (también conocida como *Ridge Regression*). Con regularización L_2, el objetivo se convierte en minimizar:

$$\text{ECM}_\lambda(\mathbf{w}) = \frac{1}{n} \|\mathbf{y} - \mathbf{X}\mathbf{w}\|_2^2 + \lambda \|\mathbf{w}\|_2^2,$$

es decir, el error cuadrático medio, donde $\lambda > 0$ es el parámetro de regularización que penaliza coeficientes grandes para reducir el sobreajuste. La solución exacta a la función regularizada, conocida como *Ridge Regression*, es:

$$\mathbf{w}_\lambda = (\mathbf{X}^\top \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^\top \mathbf{y}.$$

Consideren distintos valores de λ, por ejemplo $\lambda = 10^{-2}\sigma_1$, y aplique gradiente descendiente a la nueva función objetivo ECM_λ, comparando con las soluciones obtenidas mediante SVD y gradiente descendiente en el problema sin regularizar.