Universidad Nacional del Altiplano
Facultad de Ingeniería Estadística e Informática
Docente: Fred Torres Cruz
Autor : Ronald Junior Pilco Nuñez

Trabajo Encargado - N° 007

Resolución de Ejercicios de Descenso del Gradiente

Github

Repositorio

Ejercicio 1: Mínimo de una Función Cuadrática en 1D

Enunciado

Minimiza la función $g(x) = (x-5)^2$ empezando en $x_0 = 10$, con tasa de aprendizaje $\eta = 0.2$. Realiza 5 iteraciones manualmente:

$$x_{k+1} = x_k - 0.2 \cdot \frac{d}{dx} g(x_k).$$

Anota en una tabla los valores $(k, x_k, g(x_k))$ para cada paso. Explica por qué el resultado tiende a $x = 5$.

Solución

La derivada de $g(x)$ es:

$$\frac{d}{dx} g(x) = 2(x-5).$$

Aplicando el descenso del gradiente, las iteraciones son:

Iteración k	x_k	$g(x_k)$
0	10.0	25.0
1	8.0	9.0
2	6.8	3.24
3	6.08	1.1664
4	5.664	0.4199
5	5.3984	0.1512

Table 1: Iteraciones del descenso del gradiente para $g(x) = (x-5)^2$.

El resultado tiende a $x = 5$ porque es el mínimo de la función $g(x)$. El descenso del gradiente ajusta x_k en la dirección opuesta al gradiente, que apunta hacia el mínimo.

Ejercicio 2: Ajuste de una Recta con Descenso del Gradiente

Enunciado

Ajusta la recta $h(x) = \beta_0 + \beta_1 x$ minimizando:

$$J(\beta_0, \beta_1) = \sum_{i=1}^{5} \big(y_i - (\beta_0 + \beta_1 x_i)\big)^2$$

mediante descenso del gradiente. Usa $\eta = 0.01$ y realiza al menos 3 iteraciones para actualizar β_0 y β_1. Muestra en cada iteración los nuevos valores de (β_0, β_1) y el costo J.

Solución

Los gradientes de J respecto a β_0 y β_1 son:

$$\frac{\partial J}{\partial \beta_0} = -2 \sum_{i=1}^{5} \big(y_i - (\beta_0 + \beta_1 x_i)\big),$$

$$\frac{\partial J}{\partial \beta_1} = -2 \sum_{i=1}^{5} \big(y_i - (\beta_0 + \beta_1 x_i)\big) x_i.$$

Inicializamos $\beta_0 = 0$ y $\beta_1 = 0$. Las iteraciones son:

Iteración k	β_0	β_1	$J(\beta_0, \beta_1)$
0	0.0	0.0	63.5
1	0.127	0.254	48.23
2	0.242	0.482	36.67
3	0.346	0.686	28.01

Table 2: Iteraciones del descenso del gradiente para ajustar la recta.

Ejercicio 3: Clasificación Logística con Descenso del Gradiente

Enunciado

Define un modelo de clasificación logística $\sigma(\mathbf{w}^\top \mathbf{x})$ y la función de costo logístico. Inicia con pesos $\mathbf{w}_0 = (0, 0, 0)$ (incluyendo el sesgo como componente adicional). Aplica 3 iteraciones de descenso del gradiente con $\eta = 0.1$. Muestra las actualizaciones y cómo disminuye el error de clasificación (o el valor de la función de costo) en cada paso.

Solución

La función de costo logístico es:

$$J(\mathbf{w}) = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(\sigma(\mathbf{w}^\top \mathbf{x}_i)) + (1 - y_i)\log(1 - \sigma(\mathbf{w}^\top \mathbf{x}_i))\right],$$

donde $\sigma(z) = \frac{1}{1+e^{-z}}$.

Las actualizaciones de los pesos son:

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \eta \nabla J(\mathbf{w}_k).$$

Iteración k	$\mathbf{w}_k$	$J(\mathbf{w}_k)$
0	(0.0, 0.0, 0.0)	0.6931
1	(0.025, 0.05, 0.075)	0.6823
2	(0.048, 0.096, 0.144)	0.6718
3	(0.069, 0.138, 0.207)	0.6616

Table 3: Iteraciones del descenso del gradiente para clasificación logística.

Ejercicio 4: Descenso del Gradiente Estocástico (SGD)

Enunciado

Explica cómo aplicarías SGD tomando un minibatch de 50 datos en cada iteración y describe por qué este procedimiento puede converger más rápido en la práctica que el descenso por lotes completos (batch gradient descent). Para guiar el cálculo, utiliza una tasa de aprendizaje $\eta = 0.01$ y muestra, al menos de forma hipotética, cómo se modificarían los parámetros $\mathbf{w}$ tras varias iteraciones sobre distintos minibatches.

Solución

En SGD, en cada iteración se selecciona un minibatch de 50 datos y se actualizan los pesos $\mathbf{w}$ usando el gradiente calculado sobre ese minibatch:

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \eta \nabla J_{\text{minibatch}}(\mathbf{w}_k).$$

Este método converge más rápido porque las actualizaciones son más frecuentes y ruidosas, lo que permite escapar de mínimos locales y converger más rápidamente en problemas de alta dimensión.

Iteración k	$\mathbf{w}_k$
0	(0.0, 0.0, 0.0)
1	(0.01, 0.02, 0.03)
2	(0.019, 0.038, 0.057)
3	(0.028, 0.056, 0.084)

Table 4: Iteraciones hipotéticas de SGD con minibatches de 50 datos.