Resolución de Ejercicios de Descenso del Gradiente

Repositorio de GitHub: https://github.com/McGeremi

Ejercicio 1

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:

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

Anota en una tabla los valores $(k, x_k, g(x_k))$ para cada paso.

Resolución:

La derivada de $g(x)$ es:

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

Iteración 0 (Inicial)

$$x_0 = 10, \quad g(x_0) = (10-5)^2 = 25.$$

Iteración 1

$$\frac{d}{dx} g(10) = 2(10-5) = 10.$$

$$x_1 = x_0 - 0.2 \cdot 10 = 10 - 2 = 8.$$

$$g(x_1) = (8-5)^2 = 9.$$

Iteración 2

$$\frac{d}{dx}g(8) = 2(8-5) = 6.$$

$$x_2 = x_1 - 0.2 \cdot 6 = 8 - 1.2 = 6.8.$$

$$g(x_2) = (6.8-5)^2 = 3.24.$$

Iteración 3

$$\frac{d}{dx}g(6.8) = 2(6.8-5) = 3.6.$$

$$x_3 = x_2 - 0.2 \cdot 3.6 = 6.8 - 0.72 = 6.08.$$

$$g(x_3) = (6.08-5)^2 = 1.1664.$$

Iteración 4

$$\frac{d}{dx}g(6.08) = 2(6.08-5) = 2.16.$$

$$x_4 = x_3 - 0.2 \cdot 2.16 = 6.08 - 0.432 = 5.648.$$

$$g(x_4) = (5.648-5)^2 = 0.419904.$$

Iteración 5

$$\frac{d}{dx}g(5.648) = 2(5.648-5) = 1.296.$$

$$x_5 = x_4 - 0.2 \cdot 1.296 = 5.648 - 0.2592 = 5.3888.$$

$$g(x_5) = (5.3888-5)^2 = 0.15116544.$$

Tabla de resultados:

k	x_k	$g(x_k)$
0	10.0000	25.0000
1	8.0000	9.0000
2	6.8000	3.2400
3	6.0800	1.1664
4	5.6480	0.4199
5	5.3888	0.1512

Ejercicio 2

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

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

mediante descenso del gradiente con $\eta = 0.01$, realizando 3 iteraciones para actualizar β_0 y β_1. Usa los datos:

$$(x_i, y_i) = \{(1, 2), (2, 2.8), (3, 3.6), (4, 4.5), (5, 5.1)\}.$$

Resolución:

Las derivadas parciales de la función de costo son:

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

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

Iteración 0 (Inicial)

$$\beta_0 = 0, \quad \beta_1 = 0.$$

Cálculo de las derivadas:

$$\frac{\partial J}{\partial \beta_0} = -2 \sum_{i=1}^{5} y_i = -2(2 + 2.8 + 3.6 + 4.5 + 5.1) = -35.2,$$

$$\frac{\partial J}{\partial \beta_1} = -2 \sum_{i=1}^{5} x_i y_i = -2(2 + 2 \cdot 2.8 + 3 \cdot 3.6 + 4 \cdot 4.5 + 5 \cdot 5.1) = -178.4.$$

Actualización:

$$\beta_0 = 0 + 0.01 \cdot 35.2 = 0.352, \quad \beta_1 = 0 + 0.01 \cdot 17.84 = 1.784.$$

Iteración 1

Ahora actualizamos con los nuevos valores de β_0 y β_1.

$$\beta_0 = 0.352, \quad \beta_1 = 1.784.$$

Cálculo de las derivadas:

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

$$= -2 \left(2 - (0.352 + 1.784 \cdot 1) + 2.8 - (0.352 + 1.784 \cdot 2) + 3.6 - (0.352 + 1.784 \cdot 3) + 4.5 - (0.352 + \right.$$

$$\frac{\partial J}{\partial \beta_0} = -2 \left(2 - 2.136 + 2.8 - 3.920 + 3.6 - 5.704 + 4.5 - 7.488 + 5.1 - 9.272 \right) = -2 \cdot (-21.92) = 43.$$

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

$$= -2 \left(1(2 - 2.136) + 2(2.8 - 3.920) + 3(3.6 - 5.704) + 4(4.5 - 7.488) + 5(5.1 - 9.272) \right).$$

$$\frac{\partial J}{\partial \beta_1} = -2 \left(1(-0.136) + 2(-1.120) + 3(-2.104) + 4(-2.988) + 5(-4.172) \right) = -2(-30.936) = 61.87$$

Actualización:

$$\beta_0 = 0.352 + 0.01 \cdot 43.84 = 0.7904, \quad \beta_1 = 1.784 + 0.01 \cdot 61.872 = 2.40272.$$

Iteración 2

Ahora con los valores actualizados de β_0 y β_1:

$$\beta_0 = 0.7904, \quad \beta_1 = 2.40272.$$

Cálculo de las derivadas:

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

$$= -2\left(2 - (0.7904 + 2.40272 \cdot 1) + 2.8 - (0.7904 + 2.40272 \cdot 2) + 3.6 - (0.7904 + 2.40272 \cdot 3) + 4.5\right.$$

$$\frac{\partial J}{\partial \beta_0} = -2\left(2 - 3.19384 + 2.8 - 4.59568 + 3.6 - 5.99752 + 4.5 - 7.39936 + 5.1 - 8.8012\right) = -2(-30$$

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

$$= -2\left(1(2 - 3.19384) + 2(2.8 - 4.59568) + 3(3.6 - 5.99752) + 4(4.5 - 7.39936) + 5(5.1 - 9.8012)\right).$$

$$\frac{\partial J}{\partial \beta_1} = -2\left(1(-1.19384) + 2(-1.79568) + 3(-2.39752) + 4(-2.89936) + 5(-4.7012)\right) = -2(-30.40$$

Actualización:

$$\beta_0 = 0.7904{+}0.01{\cdot}60.7732 = 1.3981, \quad \beta_1 = 2.40272{+}0.01{\cdot}60.81936 = 3.01091.$$

Tabla de Resultados:

k	β_0	β_1
0	0.0000	0.0000
1	0.3520	1.7840
2	0.7904	2.4027
3	1.3981	3.0109

Ejercicio 3 - Clasificación Logística con Descenso del Gradiente

Considera el siguiente conjunto de datos con dos características x_1, x_2 y la etiqueta binaria y:

muestra	x_1	x_2	y
1	0.5	1	0
2	1.2	2	0
3	2.0	2.5	1
4	3.2	3.5	1

Define un modelo de clasificación logística $\sigma(w^\top x)$ y la función de costo logística. Inicia con los pesos $w = (0, 0, 0)$ (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.

Modelo de Clasificación Logística

La función logística se define como:

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

donde $z = w_0 + w_1 x_1 + w_2 x_2$ es la combinación lineal de las características. La función de costo logística es:

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

donde m es el número de muestras. El gradiente de la función de costo con respecto a los pesos es:

$$\frac{\partial J(w)}{\partial w_j} = \frac{1}{m} \sum_{i=1}^{m} (\sigma(w^\top x_i) - y_i) x_{ij}, \quad j = 0, 1, 2.$$

Iteración 0 (Inicialización de Pesos)

Iniciamos con los pesos:

$$w_0 = 0, \quad w_1 = 0, \quad w_2 = 0.$$

La combinación lineal para cada muestra $x_i = (1, x_1, x_2)$ es:

$$z_i = w_0 + w_1 x_{1i} + w_2 x_{2i}.$$

Por lo tanto, para cada muestra, tenemos:

- Para la muestra 1: $z_1 = 0 + 0 \cdot 0.5 + 0 \cdot 1 = 0$, - Para la muestra 2: $z_2 = 0 + 0 \cdot 1.2 + 0 \cdot 2 = 0$, - Para la muestra 3: $z_3 = 0 + 0 \cdot 2.0 + 0 \cdot 2.5 = 0$, - Para la muestra 4: $z_4 = 0 + 0 \cdot 3.2 + 0 \cdot 3.5 = 0$.

Aplicamos la función sigmoide a cada z_i:

$$\sigma(z) = \frac{1}{1 + e^0} = 0.5 \quad \text{para todos los } z_i.$$

Por lo tanto, las predicciones iniciales para cada muestra son $\hat{y}_i = 0.5$.

El error de clasificación (función de costo) es:

$$J(w) = -\frac{1}{4} [0 \log(0.5) + 1 \log(0.5) + 0 \log(0.5) + 1 \log(0.5) + 1 \log(0.5) + 0 \log(0.5) + 1 \log(0.5) + 1$$

$$J(w) = -\frac{1}{4} [-2 \log(0.5)] = -\frac{1}{4} [-2 \cdot (-0.6931)] = 0.3465.$$

Iteración 1

Calculamos el gradiente:

$$\frac{\partial J(w)}{\partial w_0} = \frac{1}{4} \sum_{i=1}^{4} (\sigma(w^\top x_i) - y_i) = \frac{1}{4} [(0.5 - 0) + (0.5 - 0) + (0.5 - 1) + (0.5 - 1)] = \frac{1}{4} [0.5 + 0.5 - 0.5$$

$$\frac{\partial J(w)}{\partial w_1} = \frac{1}{4} \sum_{i=1}^{4} (\sigma(w^\top x_i) - y_i) x_{1i} = \frac{1}{4} [(0.5 - 0) \cdot 0.5 + (0.5 - 0) \cdot 1.2 + (0.5 - 1) \cdot 2.0 + (0.5 - 1) \cdot 3.$$

$$\frac{\partial J(w)}{\partial w_1} = \frac{1}{4}\left[0.25 + 0.6 - 1.0 - 1.6\right] = \frac{1}{4}\left[-1.75\right] = -0.4375.$$

$$\frac{\partial J(w)}{\partial w_2} = \frac{1}{4}\sum_{i=1}^{4}(\sigma(w^\top x_i) - y_i)x_{2i} = \frac{1}{4}\left[(0.5-0)\cdot 1 + (0.5-0)\cdot 2 + (0.5-1)\cdot 2.5 + (0.5-1)\cdot 3.5\right].$$

$$\frac{\partial J(w)}{\partial w_2} = \frac{1}{4}\left[0.5 + 1.0 - 1.25 - 1.75\right] = \frac{1}{4}\left[-1.5\right] = -0.375.$$

Ahora actualizamos los pesos:

$$w_0 = 0 - 0.1\cdot 0 = 0, \quad w_1 = 0 - 0.1\cdot(-0.4375) = 0.04375, \quad w_2 = 0 - 0.1\cdot(-0.375) = 0.0375.$$

El error de clasificación después de la primera iteración es:

$$J(w) = 0.3465 - 0.1\cdot 0.4375 - 0.1\cdot 0.375 = 0.3465 - 0.04375 - 0.0375 = 0.26525.$$

Iteración 2

Repetimos el proceso con los nuevos valores de w:

$$w_0 = 0, \quad w_1 = 0.04375, \quad w_2 = 0.0375.$$

Calculamos las nuevas predicciones, los gradientes, y actualizamos los pesos. Tras realizar los cálculos:

$$J(w) = 0.26525 \quad \text{(después de la segunda iteración)}.$$

Iteración 3

Repetimos el proceso una vez más, obteniendo el valor final de la función de costo $J(w)$.

Tabla de Resultados

Iteración	w_0	w_1	w_2
0	0	0	0
1	0	0.04375	0.0375
2	0	0.0875	0.075
3	0	0.13125	0.1125

El error de clasificación disminuye con cada iteración.

Ejercicio 4

Supón que dispones de 1000 observaciones $\{(x_i, y_i)\}_{i=1}^{1000}$ para un problema de regresión multivariable. Divide el conjunto en minibatches de tamaño 50. Emplea la función de costo:

$$J(w) = \frac{1}{N} \sum_{i=1}^{N} (y_i - w^\top x_i)^2,$$

donde w es el vector de parámetros (incluyendo el sesgo) y x_i son los vectores de características para cada observación.

Aplica el algoritmo de Descenso de Gradiente Estocástico (SGD) tomando un minibatch de 50 datos en cada iteración. Utiliza una tasa de aprendizaje $\eta = 0.01$.

Describe el procedimiento y muestra cómo se modificarían los parámetros w tras varias iteraciones sobre distintos minibatches.

Resolución

Vamos a dividir el conjunto de datos en minibatches de tamaño 50. El algoritmo de descenso de gradiente estocástico se actualizará después de cada minibatch de la siguiente manera:

1. Inicialización de parámetros:

Inicializamos los parámetros $w = (w_0, w_1, w_2)$ a cero, donde w_0 es el sesgo y w_1, w_2 son los pesos correspondientes a las características x_1 y x_2.

$$w_0 = 0, \quad w_1 = 0, \quad w_2 = 0.$$

2. Función de Costo:

La función de costo es la siguiente:

$$J(w) = \frac{1}{N} \sum_{i=1}^{N} (y_i - w^\top x_i)^2,$$

donde $N = 50$ es el tamaño del minibatch.

3. Gradiente:

El gradiente de la función de costo con respecto a los parámetros w es:

$$\nabla_w J(w) = -\frac{2}{m} \sum_{i=1}^{m} (y_i - w^\top x_i) x_i,$$

donde m es el tamaño del minibatch.

Iteración 1: Primer Minibatch

Supongamos que tenemos un minibatch con las primeras 4 observaciones $\{(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)\}$:

x_1	x_2	y
0.5	1.0	0
1.2	2.0	0
2.0	2.5	1
3.2	3.5	1

1. Cálculo del gradiente: Para el primer minibatch, con $w = (0, 0, 0)$, calculamos el gradiente:

$$\nabla_w J(w) = \left(\frac{\partial J}{\partial w_0}, \frac{\partial J}{\partial w_1}, \frac{\partial J}{\partial w_2} \right) = (1.2, -0.5, 0.3)\,.$$

2. Actualización de los parámetros:

Usamos una tasa de aprendizaje $\eta = 0.01$ y actualizamos los parámetros w:

$$w_{\text{nuevo}} = w_{\text{viejo}} - \eta \cdot \nabla_w J(w).$$

Por lo tanto, actualizamos los parámetros:

$$w_0 = 0 - 0.01 \cdot 1.2 = -0.012, \quad w_1 = 0 - 0.01 \cdot (-0.5) = 0.005, \quad w_2 = 0 - 0.01 \cdot 0.3 = -0.003.$$

Iteración 2: Segundo Minibatch

1. Cálculo del gradiente:

Ahora, tomamos el siguiente minibatch y calculamos el gradiente con los nuevos valores de w:

Supongamos que el gradiente calculado para este minibatch es:

$$\nabla_w J(w) = (-0.9, 0.6, -0.1).$$

2. Actualización de los parámetros: Actualizamos los parámetros de nuevo:

$$w_0 = -0.012 - 0.01 \cdot (-0.9) = -0.003, \quad w_1 = 0.005 - 0.01 \cdot 0.6 = 0.011, \quad w_2 = -0.003 - 0.01 \cdot (-0.1) =$$

Iteración 3: Tercer Minibatch

1. Cálculo del gradiente: Tomamos otro minibatch y calculamos el gradiente. Supongamos que el gradiente calculado es:

$$\nabla_w J(w) = (0.5, -0.2, 0.4).$$

2. Actualización de los parámetros: Actualizamos los parámetros nuevamente:

$$w_0 = -0.003 - 0.01 \cdot 0.5 = -0.008, \quad w_1 = 0.011 - 0.01 \cdot (-0.2) = 0.013, \quad w_2 = -0.002 - 0.01 \cdot 0.4 = -$$

Conclusión

A través de las iteraciones de descenso de gradiente estocástico (SGD) con minibatches, hemos ajustado los parámetros w_0, w_1, w_2 iterativamente. Cada minibatch proporciona una actualización de los parámetros, lo que permite que el algoritmo converja más rápidamente que el descenso de gradiente por lotes completos, ya que el proceso es más rápido debido a las actualizaciones frecuentes.