1. Evaluate θ^1 using Stochastic Gradient Descent

We are asked to perform one step of Stochastic Gradient Descent (SGD) to find θ^1.

Step 1: Define Model, Loss, and Data

- **Model Hypothesis ($\hat{y}$):** $\hat{y} = h(x_1, x_2) = \sigma(z)$
- **Linear Input (z):** $z = b + w_1 x_1 + w_2 x_2$
- **Parameters (θ):** $\theta = (b, w_1, w_2)$
- **Loss Function (L):** For a regression problem (house price), we use the **Squared Error Loss**.

$$L(\theta) = \frac{1}{2}(\hat{y} - y)^2 = \frac{1}{2}(\sigma(z) - y)^2$$

- **Initial Parameters (θ^0):** $\theta^0 = (b^0, w_1^0, w_2^0) = (4, 5, 6)$
- **Data Point (x, y):** $(x_1, x_2, y) = (1, 2, 3)$

Step 2: Define the SGD Update Rule

The SGD update rule for the new parameters $\theta^1 = (b^1, w_1^1, w_2^1)$ is:

$$\theta^1 = \theta^0 - \alpha \nabla L(\theta^0)$$

where α is the learning rate (which is not given, so we keep it as a symbol). We need to find the gradient vector $\nabla L = \left[\frac{\partial L}{\partial b}, \frac{\partial L}{\partial w_1}, \frac{\partial L}{\partial w_2}\right]$.

Step 3: Calculate the Gradients (Chain Rule)

We find the partial derivative of L with respect to each parameter.

$$\frac{\partial L}{\partial \theta_j} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial z}{\partial \theta_j}$$

Let's find the common components first:

- $\frac{\partial L}{\partial \hat{y}} = \frac{\partial}{\partial \hat{y}}\left[\frac{1}{2}(\hat{y} - y)^2\right] = (\hat{y} - y)$
- $\frac{\partial \hat{y}}{\partial z} = \frac{d\sigma(z)}{dz} = \sigma(z)(1 - \sigma(z)) = \hat{y}(1 - \hat{y})$

Now, we find the specific gradients:

$$\begin{aligned}
\frac{\partial L}{\partial b} &= \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial z}{\partial b} = (\hat{y} - y) \cdot \hat{y}(1 - \hat{y}) \cdot 1 \\
\frac{\partial L}{\partial w_1} &= \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial z}{\partial w_1} = (\hat{y} - y) \cdot \hat{y}(1 - \hat{y}) \cdot x_1 \\
\frac{\partial L}{\partial w_2} &= \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial z} \cdot \frac{\partial z}{\partial w_2} = (\hat{y} - y) \cdot \hat{y}(1 - \hat{y}) \cdot x_2
\end{aligned}$$

Step 4: Substitute Numbers to find θ^1

As per the instruction, we just write the expression and substitute the numbers, without simplifying.

First, we express the prediction $\hat{y}^0$ made with the initial parameters θ^0 and data x:

$$z^0 = b^0 + w_1^0 x_1 + w_2^0 x_2 = 4 + 5 \cdot 1 + 6 \cdot 2$$

$$\hat{y}^0 = \sigma(z^0) = \sigma(4 + 5 \cdot 1 + 6 \cdot 2)$$

Now we write the update equation for each parameter, substituting θ^0, $y = 3$, $x_1 = 1$, and $x_2 = 2$.

For b^1:

$$b^1 = b^0 - \alpha \cdot \left.\frac{\partial L}{\partial b}\right|_{\theta^0,x,y}$$

$$b^1 = 4 - \alpha \cdot [(\sigma(4 + 5 \cdot 1 + 6 \cdot 2) - 3) \cdot \sigma(4 + 5 \cdot 1 + 6 \cdot 2) \cdot (1 - \sigma(4 + 5 \cdot 1 + 6 \cdot 2)) \cdot 1]$$

For w_1^1:

$$w_1^1 = w_1^0 - \alpha \cdot \left.\frac{\partial L}{\partial w_1}\right|_{\theta^0,x,y}$$

$$w_1^1 = 5 - \alpha \cdot [(\sigma(4 + 5 \cdot 1 + 6 \cdot 2) - 3) \cdot \sigma(4 + 5 \cdot 1 + 6 \cdot 2) \cdot (1 - \sigma(4 + 5 \cdot 1 + 6 \cdot 2)) \cdot 1]$$

For w_2^1:

$$w_2^1 = w_2^0 - \alpha \cdot \left.\frac{\partial L}{\partial w_2}\right|_{\theta^0,x,y}$$

$$w_2^1 = 6 - \alpha \cdot [(\sigma(4 + 5 \cdot 1 + 6 \cdot 2) - 3) \cdot \sigma(4 + 5 \cdot 1 + 6 \cdot 2) \cdot (1 - \sigma(4 + 5 \cdot 1 + 6 \cdot 2)) \cdot 2]$$

Consolidated Result for θ^1: To write the final expression for θ^1 more concisely, let's define a common term G (the gradient factor):

$$G = (\sigma(4 + 5 \cdot 1 + 6 \cdot 2) - 3) \cdot \sigma(4 + 5 \cdot 1 + 6 \cdot 2) \cdot (1 - \sigma(4 + 5 \cdot 1 + 6 \cdot 2))$$

The updated parameter vector $\theta^1 = (b^1, w_1^1, w_2^1)$ can then be written as:

$$\theta^1 = (4 - \alpha G, \quad 5 - \alpha G, \quad 6 - 2\alpha G)$$

Or, using the original vector notation:

$$\theta^1 = (4, 5, 6) - \alpha G \cdot (1, 1, 2)$$