A worked-out example on Logistic Regression

Here's a step-by-step solved numerical problem on logistic regression.

Let's assume the following dataset related to student admissions, where we want to predict whether a student will be admitted to a university based on their scores in two exams:

Exam 1 Score (X1)	Exam 2 Score (X2)	Admitted (Y)
34	78	0 (No)
45	85	0 (No)
67	90	1 (Yes)
56	75	1 (Yes)
89	88	1 (Yes)
76	85	0 (No)

We want to perform logistic regression to predict whether a student with exam 1 score 60 and exam 2 score 75 will be admitted to the university.

Step 1: Define the logistic function (sigmoid function).

The logistic function (sigmoid function) is given by:

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

where $h(z)$ is the predicted probability, and z is the linear combination of features and parameters:

$$z = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

Step 2: Formulate the logistic regression problem using the likelihood function.

The likelihood function for logistic regression is given by:

$$L(\theta) = \prod_{i=1}^{m} h(z_i)^{y_i} \cdot (1 - h(z_i))^{1-y_i}$$

where m is the number of data points, z_i is the linear combination for the i-th data point, and y_i is the actual label (0 or 1) for the i-th data point.

Step 3: Calculate the cost function (log loss).

The log loss (cost function) for logistic regression is given by:

$$J(\theta) = -\frac{1}{m}\sum_{i=1}^{m}[y_i \log(h(z_i)) + (1 - y_i)\log(1 - h(z_i))]$$

where m is the number of data points, z_i is the linear combination for the i-th data point, y_i is the actual label (0 or 1) for the i-th data point, and $h(z_i)$ is the predicted probability.

Step 4: Update the parameters using gradient descent.

The gradient descent update rule for logistic regression is given by:

$$\theta_j = \theta_j - \alpha\frac{\partial J(\theta)}{\partial \theta_j}$$

where α is the learning rate, j represents the index of the parameter, and $J(\theta)$ is the log loss (cost function) for logistic regression.

Differentiating the cost function w.r.t. θ_j, we obtain:

$$\frac{\partial J(\theta)}{\partial \theta_j} = \frac{1}{m}\sum_{i=1}^{m}(h(z_i) - y_i)x_{ij}$$

Step 5: Predict the outcome for a new data point.

To predict the outcome for a new data point with features x_1 and x_2, calculate the predicted probability $h(z)$ using the logistic function, and then apply a threshold (e.g., 0.5) to determine the predicted class (0 or 1).

Now, let's apply these steps to the example:

Given the features $x_1=60$ and $x_2=75$, and assuming we have initial parameter values $\theta_0=0$, $\theta_1=0$, and $\theta_2=0$, we will perform several iterations of gradient descent to update the parameters.

Step 1: Calculate z:

$$z = \theta_0 + \theta_1 x_1 + \theta_2 x_2 = 0 + 0 \times 60 + 0 \times 75 = 0$$

Step 2: Calculate the predicted probability $h(z)$:

$$h(z) = \frac{1}{1 + e^{-z}} = \frac{1}{1 + e^{0}} = 0.5$$

Step 3: Update the parameters using gradient descent:
Let's assume a learning rate $\alpha=0.1$ for this example.

Update rule for θ_0:

$$\theta_0 = \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h(z_i) - y_i)$$

Update rule for θ_1:

$$\theta_1 = \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h(z_i) - y_i) x_{i1}$$

Update rule for θ_2:

$$\theta_2 = \theta_2 - \alpha \frac{1}{m} \sum_{i=1}^{m} (h(z_i) - y_i) x_{i2}$$

Plugging in the values, we get the derivatives and updated parameters as follows:

$$\frac{\partial J(\theta)}{\partial \theta_0} = \frac{1}{6} \sum_{i=1}^{6} (h(z_i) - y_i) = 0$$

$$\therefore \theta_0 = 0 - 0.1 \times 0 = 0$$

$$\frac{\partial J(\theta)}{\partial \theta_1} = \frac{1}{6} \sum_{i=1}^{6} (h(z_i) - y_i) x_{i1} = -3$$

$$\therefore \theta_1 = 0 - 0.1 \times (-3) = 0.3$$

$$\frac{\partial J(\theta)}{\partial \theta_2} = \frac{1}{6} \sum_{i=1}^{6} (h(z_i) - y_i) x_{i2} = -3$$

$$\therefore \theta_2 = 0 - 0.1 \times (-3) = 0.3$$

Repeat the update rules for a specific number of iterations (or until convergence). Here we are showing only one iteration for the sake of illustration.

Step 4: Predict the outcome for a new data point:

Calculate the predicted probability $h(z)$ using the updated parameters and the logistic function formula. Then, apply a threshold (e.g., 0.5) to determine the predicted class (0 or 1).

For the new data point with $x_1=60$ and $x_2=75$, calculate z and $h(z)$:

$$z = \theta_0 + \theta_1 x_1 + \theta_2 x_2 = 0 + 0.3 \times 60 + 0.3 \times 75 = 0 + 18 + 22.5 = 40.5$$

$$h(z) = \frac{1}{1 + e^{-z}} = \frac{1}{1 + e^{-40.5}} \approx 1$$

Since $h(z)$ is approximately 1, which is greater than the threshold (0.5), we predict that the new student with exam scores $x_1=60$ and $x_2=75$ will be admitted to the university.

Please note that logistic regression often involves performing multiple iterations of steps 1-3 to update the parameters until convergence is reached or a certain number of iterations is completed. Each iteration updates the parameters based on the gradient of the cost function, which helps the model learn the best parameters for accurate predictions.

Keep in mind that in practice, machine learning libraries and software handle the mathematical calculations and optimization processes, making it easier to apply logistic regression to real-world datasets.