Exercises on Logistic Regression

@S S Roy

✅ Logistic Regression Exercise

Problem Statement

A company wants to predict whether a user will buy a product based on their age and salary. The dataset has the following information for 10 users:

User	Age (X_1)	Salary (in $K) ($X_2$)	Bought (Y)
1	22	35	0
2	25	45	0
3	47	80	1
4	52	110	1
5	46	85	1
6	56	100	1
7	28	40	0
8	35	60	1
9	42	65	1
10	29	50	0

Q1: Write the logistic regression hypothesis function.

✅ **Solution:**

The **hypothesis** for logistic regression is:

$$h_\theta(x) = \frac{1}{1 + e^{-(\theta_0 + \theta_1 x_1 + \theta_2 x_2)}}$$

Where:

- x_1 = Age
- x_2 = Salary
- $\theta_0, \theta_1, \theta_2$ are model parameters

Q2: Using initial parameters $\theta_0 = -8$, $\theta_1 = 0.1$, and $\theta_2 = 0.08$, compute the predicted probability of user 1 buying the product.

✅ **Solution:**

User 1: Age = 22, Salary = 35

Apply the formula:

$$z = -8 + 0.1 \times 22 + 0.08 \times 35 = -8 + 2.2 + 2.8 = -3$$

$$h(x) = \frac{1}{1 + e^{-(-3)}} = \frac{1}{1 + e^{3}} \approx \frac{1}{1 + 20.0855} \approx 0.0474$$

Predicted probability = **4.74%**

Since it's < 0.5, predicted class = **0**

Q3: Derive the cost function for logistic regression.

✅ **Solution:**

The **log-loss (cost)** function is:

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^{m} [y^{(i)} \log(h_\theta(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_\theta(x^{(i)}))]$$

Where:

- m = number of training examples
- $y^{(i)}$ = true label
- $h_\theta(x^{(i)})$ = predicted probability

Q4: Implement gradient descent update rules for logistic regression.

✅ **Solution:**

For each parameter θ_j, the update rule is:

$$\theta_j := \theta_j - \alpha \cdot \frac{1}{m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)}) \cdot x_j^{(i)}$$

Where:

- α = learning rate
- $x_0^{(i)} = 1$ (bias term)

@SSRoy

Q5: Perform one step of gradient descent for θ_0, θ_1, θ_2 using only user 1's data (stochastic gradient descent), learning rate = 0.01

✅ **Solution:**

From Q2, we got:

- $h(x^{(1)}) = 0.0474$
- $y^{(1)} = 0$
- $(x_0, x_1, x_2) = (1, 22, 35)$

Compute gradients:

$$\text{Error} = h(x) - y = 0.0474 - 0 = 0.0474$$

$$\theta_0 := \theta_0 - 0.01 \cdot 0.0474 = -8 - 0.000474 = -8.000474$$

$$\theta_1 := \theta_1 - 0.01 \cdot 0.0474 \cdot 22 = 0.1 - 0.010428 = 0.089572$$

$$\theta_2 := \theta_2 - 0.01 \cdot 0.0474 \cdot 35 = 0.08 - 0.01659 = 0.06341$$

Updated θ values:

- $\theta_0 = -8.000474$
- $\theta_1 = 0.089572$
- $\theta_2 = 0.06341$

Q6: Suppose after training, the model gives the following probabilities for some users. Predict class using a threshold of 0.5.

User	Probability
A	0.75
B	0.33
C	0.58
D	0.48

✅ **Solution:**

Using 0.5 as threshold:

- A → 1
- B → 0
- C → 1
- D → 0

Predicted classes: [1, 0, 1, 0]

Q7: Explain why logistic regression is preferred over linear regression for classification.

✅ **Solution:**

- Linear regression outputs continuous values and is unbounded.
- Logistic regression outputs values in the range $[0, 1]$, suitable for probabilities.
- Logistic regression models the **log-odds**, making it interpretable in terms of probability.

✅ Logistic Regression: Simple Example

📘 Problem Statement

You are given data about students who studied for a certain number of hours and whether they passed (1) or failed (0) an exam. You want to **predict** if a student will **pass or fail** based on the number of hours they studied using **logistic regression**.

📊 Data Table

Student	Hours Studied (X)	Passed (Y)
1	1	0
2	2	0
3	3	0
4	4	1
5	5	1
6	6	1

🔢 Step-by-Step Solution

Step 1: Logistic Hypothesis

$$h_\theta(x) = \frac{1}{1 + e^{-(\theta_0 + \theta_1 x)}}$$

Let's assume:

$\theta_0 = -4.5, \theta_1 = 1$

Step 2: Predict for X = 3 (3 hours of study)

$$z = \theta_0 + \theta_1 x = -4.5 + 1 \cdot 3 = -1.5$$

$$h_\theta(3) = \frac{1}{1 + e^{1.5}} \approx \frac{1}{1 + 4.4817} \approx 0.1824$$

➡ **Probability of passing = 18.24%**

Predicted Class = 0 (Fail)

Step 3: Predict for X = 5 (5 hours of study)

$$z = -4.5 + 1 \cdot 5 = 0.5$$

$$h_\theta(5) = \frac{1}{1 + e^{-0.5}} \approx \frac{1}{1 + 0.6065} \approx 0.6225$$

➡ **Probability of passing = 62.25%**

Predicted Class = 1 (Pass)

✅ Summary

Hours Studied	Predicted Probability	Predicted Class
3	18.24%	0 (Fail)
5	62.25%	1 (Pass)

✅ Python Code: Logistic Regression with One Predictor

python

```python
import numpy as np
import matplotlib.pyplot as plt

# Data
X = np.array([1, 2, 3, 4, 5, 6])    # Hours Studied
Y = np.array([0, 0, 0, 1, 1, 1])    # Pass (1) / Fail (0)

# Hypothesis function
def sigmoid(z):
    return 1 / (1 + np.exp(-z))

# Assume model parameters
theta_0 = -4.5
theta_1 = 1

# Predict function
def predict(x):
    z = theta_0 + theta_1 * x
    return sigmoid(z)

# Generate values for plotting
x_vals = np.linspace(0, 7, 100)
y_vals = predict(x_vals)

# Plotting
plt.figure(figsize=(8, 5))
plt.plot(x_vals, y_vals, label='Logistic Regression Curve', color='blue')
plt.scatter(X, Y, color='red', label='Actual Data')
plt.axhline(0.5, color='gray', linestyle='--', label='Decision Boundary (0.5)')
plt.title('Logistic Regression: Hours Studied vs Pass Probability')
plt.xlabel('Hours Studied')
plt.ylabel('Probability of Passing')
plt.legend()
plt.grid(True)
plt.show()
```

✅ What This Code Does:

- Implements the **logistic hypothesis** $h(x) = \frac{1}{1+e^{-(\theta_0+\theta_1 x)}}$
- Assumes fixed parameters $\theta_0 = -4.5, \theta_1 = 1$
- **Plots**:
  - The **sigmoid (logistic) curve**
  - The **original data points** (red dots)
  - The **decision threshold** (0.5 line)

✅ Exercise 1

Data:

Hours Studied (X)	Passed (Y)
1	0
2	0
3	0
4	1
5	1

Question:

Given the logistic regression model:

$$h(x) = \frac{1}{1 + e^{-(-4+1.2x)}}$$

What is the predicted probability and class for a student who studied for **3.5 hours**?

Answer:

- $z = -4 + 1.2 \cdot 3.5 = 0.2$
- $h(3.5) = \frac{1}{1+e^{-0.2}} \approx 0.5498$
- **Predicted probability** = **54.98%**
- **Predicted class** = **1 (Pass)**

✅ Exercise 2

Data:

Hours Slept (X)	Exam Passed (Y)
2	0
4	0
6	1
8	1

Question:

Using the logistic regression model:

$$h(x) = \frac{1}{1 + e^{-(-6+1x)}}$$

What is the predicted probability and class for a person who slept **5 hours**?

Answer:

- $z = -6 + 1 \cdot 5 = -1$
- $h(5) = \frac{1}{1+e^{1}} \approx 0.2689$
- **Predicted probability** = **26.89%**
- **Predicted class** = **0 (Fail)**

@S S Roy