Bài 3: Phân tích mô hình hồi quy logistic

(Bài tập này giúp bạn nắm bắt cách chuyển các công thức toán học sang mã lệnh Python đúng đắn)

Hướng dẫn: điền đoạn code Python đúng vào phần ô trống ở cột Mã lệnh Python.

Dữ liệu của bài tập chứa trong các tập tin ex2data1.txt và ex2data2.txt, theo cấu trúc từng bộ dữ liệu (x, y) xếp trên từng hàng. Các giá trị trên từng hàng phân tách nhau bởi dấu ‘,’.

STT	Biểu diễn Toán học	Mã lệnh Python
1	Tập dữ liệu $D = \{(x_i, y_i) \mid x_i \in R^n, y_i \in \{0,1\}, \forall i = \overline{1,m}\}$. Lấy ra $X = \{x_i \mid x_i \in R^n, \forall i = \overline{1,m}\}$ và $y\{y_i \mid y_i \in \{0,1\}, \forall i = \overline{1,m}\}$ từ tập D. ***Ghi chú***: - $X \in R^{m \times n}$ - $y \in \{0,1\}^{m \times 1}$	`def readData(filePath: str, filename: str):` `  data = np.loadtxt(os.path.join(filePath, filename), delimiter = ',')` `  X = data[:,:-1]` `  y = data[:, -1]` `  m = X.shape[0]` `  n = X.shape[1]` `  X = np.reshape(X, (m,n))` `  y = np.reshape(y, (m,1))` `  return X, y`
2	Chuyển đổi không gian của X từ R^n sang $R^{(n+1)}$, bằng cách thêm 1 cột chứa các giá trị 1 vào bên trái ma trận X. - Tạo vector (ma trận $m \times 1$) chứa các số 1, $x_0 = \begin{bmatrix} x_{01} = 1 \\ \vdots \\ x_{0m} = 1 \end{bmatrix}$ - Tạo ma trận X mới: $X = [x_0, X]$ ***Ghi chú***: lúc này $X \in R^{m \times n}$ (ngầm hiểu $n = n + 1$)	`# Tạo vector (ma trận m × 1) chứa các số 1` `x0 = np.ones((m,1))` `# Thêm cột x0 vào bên trái ma trận X` `X = np.column_stack([x0, X])`

3	Về lý thuyết, ta có: $h_w(x) = \frac{1}{1+e^{-(w0.x0+ w1x1+ \cdots + wnxn)}}$ (1) hay $h_w(x) = \frac{1}{1+e^{-w^T x}}$ (2) với: - $w \in R^{n\times 1}, x \in R^{n\times 1}$ Viết lại (1) và (2) dưới dạng phép toán ma trận $h_w(X) = \frac{1}{1+e^{-X.w}}$ (3) ***Ghi chú***: - $X \in R^{m\times n}, w \in R^{n\times 1} \rightarrow X.w \in R^{m\times 1}$ - $h_w(X) \in R^{m\times 1}$	*- Hãy lập trình Python tính giá trị **h** phương trình (3) (gợi ý: hàm np.exp() – tính số mũ)* `def sigmoid(X, w):` `  result = 1/(1 + np.exp(-np.dot(X, w)))` `  return result`
4	Hàm mất mát $J(w) = -\frac{1}{m}\sum_{i=1}^{m}[y_i.\log(h_w(x_i)) + (1-y_i).\log(1-h_w(x_i))]$ (4) Viết lại (4) theo phép toán ma trận $J(w) = -\frac{1}{m}\sum[y^T \log(h_w(X)) + (1-y)^T log(1-h_w(X))]$ (5) Ghi chú: - $y \in R^{m\times 1} \rightarrow y^T \in R^{1\times m}$ - $h_w(X) \in R^{m\times 1} \rightarrow \log(h_w(X)) \in R^{m\times 1}$ - $(1-y) \in R^{m\times 1} \rightarrow (1-y)^T \in R^{1\times m}$ - $(1-h_w(X)) \in R^{m\times 1} \rightarrow log(1-h_w(X)) \in R^{m\times 1}$ - $J(w) \in R$	`def loss(X, y, w):` `  m = y.shape[0]` `  result=(-1/m)*np.sum(np.dot(y.T, np.log(sigmoid(X, w))) + np.dot((1 - y).T, np.log(1 - sigmoid(X, w))))` `  return result`
5	Theo lý thuyết, công thức tính đạo hàm riêng là: $\frac{\partial J(w)}{\partial w_i} = \frac{1}{m}\sum_{i=1}^{m}(h_w(x_i) - y_i).x_i$ (6) Viết lại (6) theo phép toán ma trận: $\nabla J(w) = \frac{1}{m}\sum X^T(h_w(X) - y)$ (7) Ghi chú: - $h_w(X) \in R^{m\times 1}, y \in R^{m\times 1} \rightarrow (h_w(X) - y) \in R^{m\times 1}$ - $X \in R^{m\times n} \rightarrow X^T \in R^{n\times m}$ - *$\nabla J(w) \in R^{n\times 1}$ chính là cùng kích thước ma trận w*	*-Hãy viết mã lệnh tính $\nabla J(w)$ theo công thức (7)* `def gradient(X, y, w):` `  m = X.shape[0]` `  result = (1/m)*np.dot(X.T, sigmoid(X, w) - y)` `  return result`

<table>
<tr>
<td>6</td>
<td>Theo thuật toán Gradient descent, các trọng số được cập nhật theo công thức

$w_i = w_i - \alpha \frac{\partial J(w)}{\partial w_i}, \forall i$ (8)

Viết lại (8) dưới dạng phép toán ma trận:

$w = w - \alpha . \nabla J(w)$ (9)

Ghi chú:
- $w \in R^{n\times 1}, . \nabla J(w) \in R^{n\times 1} \rightarrow w - \alpha . \nabla J(w) \in R^{n\times 1}$</td>
<td>- Hãy viết mã lệnh tính w theo công thức (9)


```
def gradientDescent(X, y, w, alpha, n_iters):
    w_optimal = w.copy()
    J_history = []
    for i in range(n_iters):
        w_optimal = w_optimal -
 alpha*gradient(X, y, w_optimal)
        J_history.append(loss(X, y, w_optimal))
    return w_optimal, J_history
```

</td>
</tr>
<tr>
<td>7</td>
<td colspan="2">Hãy hoàn thành chương trình xây dựng mô hình Logistic có xét đến trường hợp chuẩn hóa dữ liệu và không chuẩn hóa dữ liệu.Ghi chú: Áp dụng 2 phương phasp chuẩn hóa dữ liệu:
- Mean normalization
- Max-Min normalization</td>
</tr>
</table>