

Lab assignment 2 – Logistic regression

1. Yêu cầu

- Cài đặt **hồi quy logistic** để dự đoán microchip của một nhà máy có đủ điều kiện để được bán ra thị trường hay không.
- Dữ liệu thô bao gồm 3 cột: cột 1 và 2 là các đặc trưng, cột 3 là nhãn.
- Dữ liệu huấn luyện là dữ liệu thô sau khi ánh xạ sang miền dữ liệu mới gồm 28 chiều. Hàm `map_feature` được cung cấp trong file **map_feature.py** sẽ thực hiện việc này.
- Cài đặt các hàm phụ trợ sau để thực hiện việc huấn luyện và dự đoán:
 - `compute_cost`: tính toán chi phí của mô hình trên tập dữ liệu (công thức hàm chi phí được cung cấp trong phần 3).
 - `compute_gradient`: tính vector gradient của hàm chi phí (công thức tính vector gradient được cung cấp trong phần 3).
 - `gradient_descent`: tính gradient descent.
 - `predict`: dự đoán một tập các microchip có đủ điều kiện để bán ra thị trường hay không (nếu muốn dự đoán cho 1 microchip thì phải truyền vào một mảng gồm 1 phần tử).
 - `evaluate`: đánh giá kết quả dự đoán của mô hình dựa trên các độ đo: **accuracy**, **precision**, **recall** và **F1-score** (tương tự phương thức **classification_report** của **scikit-learn**).
- Tiếp theo đó, sinh viên cần sử dụng các hàm phụ trợ để cài đặt phần chương trình chính bao gồm các yêu cầu sau:
 - Đọc cấu hình huấn luyện từ tập tin **config.json**.
 - Huấn luyện với dữ liệu được cung cấp từ tập tin **training_data.txt**.
 - Lưu mô hình huấn luyện được vào tập tin **model.json**.
 - Dự đoán và đánh giá kết quả huấn luyện trên tập dữ liệu huấn luyện, lưu kết quả đánh giá vào tập tin **classification_report.json**.

2. Quy định nộp bài

- Đặt tất cả mã nguồn và tập tin liên quan vào thư mục có tên <MSSV>.
- Nén thư mục <MSSV> lại và nộp thông qua Moddle.

3. Các công thức

- Hàm chi phí được tính bởi công thức sau:

$$J(\theta) = \frac{1}{m} \sum_{i=1}^m [-y^{(i)} \log(h_{\theta}(x^{(i)})) - (1 - y^{(i)}) \log(1 - h_{\theta}(x^{(i)}))] + \frac{\lambda}{2m} \sum_{j=1}^n \theta_j^2.$$

- Công thức tính vector gradient:

$$\frac{\partial J(\theta)}{\partial \theta_0} = \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) x_j^{(i)} \quad \text{for } j = 0$$

$$\frac{\partial J(\theta)}{\partial \theta_j} = \left(\frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)}) x_j^{(i)} \right) + \frac{\lambda}{m} \theta_j \quad \text{for } j \geq 1$$