

Bài tập thực hành 2 - Hồi quy logistic

Yêu cầu

- Cài đặt **hồi quy logistic với regularization** để dự đoán microchip của một nhà máy có đủ điều kiện để được bán ra thị trường hay không.
- Dữ liệu thô gồm 3 cột: cột 1 và 2 là input và cột 3 là nhãn.
- Dữ liệu huấn luyện là dữ liệu thô sau khi ánh xạ sang miền dữ liệu mới gồm 28 chiều (28 đặc trưng). Hàm mapFeature đã được cung cấp sẵn để làm việc này.
 - Lưu ý: cột đầu tiên của dữ liệu ánh xạ có giá trị 1, trong quá trình huấn luyện chúng ta không cần thêm '1' nữa.
- Cài đặt các hàm sau để thực hiện việc huấn luyện và dự đoán
 - sigmoid: hàm sigmoid trên numpy array
 - computeCost: tính chi phí của mô hình trên tập dữ liệu
 - computeGradient: tính vector gradient của hàm chi phí
 - gradientDescent: thuật toán huấn luyện
 - predict: dự đoán microchip có đủ điều kiện để bán ra thị trường hay không
 - Lưu ý: các phép toán phải được thực hiện trên vector, sử dụng thư viện numpy
- Chương trình chính:
 - Đọc cấu hình huấn luyện từ tập tin **config.json**
 - Huấn luyện, dữ liệu từ tập tin ex2data2.txt
 - Lưu mô hình vào tập tin **model.json**
 - Dự đoán và tính độ chính xác trên tập dữ liệu huấn luyện, lưu kết quả vào **accuracy.json**
 - Lưu ý: đặt đúng tên 3 tập tin json trên
- Dữ liệu: dữ liệu huấn luyện và các tập tin json tải tại link sau:
<https://drive.google.com/drive/folders/1TYXLl8k8lt1baRaV9tpW5ppyXtpVZvEX>

Quy định nộp bài

- Đặt tất cả mã nguồn và tập tin liên quan và thư mục có tên [MSSV]
- Nén thư mục thành tập tin [MSSV].zip

Hướng dẫn

- 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:

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

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