**I. Logistic Regression**

**1.1 What is Logistic Regression?**

Logistic Regression là một quá trình mô hình hóa xác xuất của một kết quả riêng biệt vơới biến đầu vào. Nó được sử dụng để dự đoán một biến phụ thuộc có giá trị rời rạc (nhị phân, đa lớp,..) dựa trên một tập hợp các biến độc lập.

# 1.2 Binary Logistic Classification

Hồi quy nhị phân được biểu diễn như là một thứ gì đó có thể nhận 2 giá trị như True/ False, Yes/ No,

0/1,..

# 1.3 Training a Binary Classifier

Đối với bài toán phân loại nhị phân, chúng ta có thể sử dụng Logistic Regression để dự đoán xác suất của một mẫu thuộc về một lớp nào đó. Nếu xác suất lớn hơn 50%, mẫu đó sẽ được dự đoán là thuộc về lớp đó, ngược lại thì không.

Để huấn luyện một mô hình Logistic Regression, chúng ta cần tính toán độ lỗi của mô hình. Để tính toán độ lỗi, chúng ta cần một hàm mất mát. Hàm mất mát thường được sử dụng trong Logistic Regression là Cross-Entropy Loss.

Về phần phương pháp tối ưu, chúng ta có thể sử dụng Gradient Descent để tối ưu hàm mất mát.

Để đánh giá mô hình, chúng ta có thể sử dụng các độ đo như Accuracy, Precision, Recall, F1-score,..

## 1.3.1 Cross-Entropy Loss

Cross-Entropy Loss được sử dụng để đo lường độ lỗi của mô hình phân loại. Nó được tính bằng cách tính toán tổng các điểm dữ liệu trong tập huấn luyện. Để tính toán Cross-Entropy Loss, chúng ta cần biết xác suất dự đoán của mô hình cho mỗi điểm dữ liệu. Để tính toán xác suất dự đoán, chúng ta cần sử dụng hàm Sigmoid.

Hàm Sigmoid được sử dụng để chuyển đổi giá trị đầu ra của mô hình thành một giá trị xác suất nằm trong khoảng từ 0 đến 1. Hàm Sigmoid được định nghĩa như sau:

Công thức Cross-Entropy Loss:

Đạo hàm của Cross-Entropy Loss:

là xác suất dự đoán của mô hình cho điểm dữ liệu thứ i.

là nhãn của điểm dữ liệu thứ i.

Trong đó:

là giá trị của thuộc tính thứ j của điểm dữ liệu thứ i. là tham số thứ j của mô hình.

## 1.3.2 Gradient Descent

Gradient Descent là một phương pháp tối ưu hóa được sử dụng để tìm ra bộ tham số tối ưu cho mô hình. Phương pháp này được sử dụng để tối ưu hàm mất mát bằng cách cập nhật các tham số của mô hình theo hướng ngược với đạo hàm của hàm mất mát.

Công thức cập nhật tham số:

là learning rate.

Trong đó:

là đạo hàm của hàm mất mát theo tham số thứ j.

Phương pháp cập nhật tham số theo Gradient Descent được thực hiện cho đến khi hàm mất mát không thay đổi nhiều hoặc đạt đến một ngưỡng nhất định.

Có 3 loại Gradient Descent:

Batch Gradient Descent: Cập nhật tham số sau khi tính toán đạo hàm trên toàn bộ tập huấn luyện.

Stochastic Gradient Descent: Cập nhật tham số sau khi tính toán đạo hàm trên một điểm dữ liệu.

Mini-batch Gradient Descent: Cập nhật tham số sau khi tính toán đạo hàm trên một batch dữ liệu.

### 1.3.2.1 Batch Gradient Descent

Batch Gradient Descent là phương pháp cập nhật tham số sau khi tính toán đạo hàm trên toàn bộ tập huấn luyện. Đây là phương pháp cập nhật tham số chậm nhất vì chúng ta phải tính toán đạo hàm trên toàn bộ tập huấn luyện trước khi cập nhật tham số.

Cách làm này có một vài hạn chế đối với cơ sở dữ liệu có vô cùng nhiều điểm.Việc phải tính toán lại đạo hàm với tất cả các điểm này sau mỗi vòng lặp trở nên cồng kềnh và không hiệu quả Công thức cập nhật tham số:

### 1.3.2.2 Stochastic Gradient Descent

Stochastic Gradient Descent là phương pháp cập nhật tham số sau khi tính toán đạo hàm trên một điểm dữ liệu. Đây là phương pháp cập nhật tham số nhanh nhất vì chúng ta chỉ phải tính toán đạo hàm trên một điểm dữ liệu trước khi cập nhật tham số.

Đối với thuật toán này, tại một điểm dữ liệu và cập nhật tham số , công thức cập nhật tham số sẽ được viết lại như sau:

là điểm dữ liệu thứ i. là tham số của mô hình.

Trong đó:

là learning rate.

là xác suất dự đoán của mô hình cho điểm dữ liệu thứ i.

### 1.3.2.3 Mini-batch Gradient Descent

Mini-batch Gradient Descent là phương pháp cập nhật tham số sau khi tính toán đạo hàm trên một batch dữ liệu. Đây là phương pháp cập nhật tham số trung bình giữa Batch Gradient Descent và Stochastic Gradient Descent.

Khác với SGD, mini-batch sử dụng một số lượng n lớn hơn 1 (nhưng vẫn nhỏ hơn tổng số dữ liệu N rất nhiều). Giống với SGD, Mini-batch Gradient Descent bắt đầu mỗi epoch bằng việc xáo trộn ngẫu nhiên dữ liệu rồi chia toàn bộ dữ liệu thành các mini-batch, mỗi mini-batch có n điểm dữ liệu (trừ mini-batch cuối có thể có ít hơn nếu N không chia hết cho n ). Mỗi lần cập nhật, thuật toán này lấy ra một mini-batch để tính toán đạo hàm rồi cập nhật. Công thức có thể viết dưới dạng:

là điểm dữ liệu thứ i. là tham số của mô hình. là learning rate.

Trong đó:

là xác suất dự đoán của mô hình cho điểm dữ liệu thứ i.

II. Programming Assignment

File code đi kèm