Báo Cáo Về Thuật Toán Perceptron Trên Tập Dữ Liệu Iris

1. Giới thiệu

Perceptron là một trong những thuật toán cơ bản trong học máy, được phát triển bởi Frank Rosenblatt vào những năm 1950. Thuật toán này là một ví dụ đơn giản của một mạng neuron nhân tạo và được sử dụng để giải quyết bài toán phân loại nhị phân (binary classification). Trong báo cáo này, chúng tôi sẽ áp dụng thuật toán Perceptron để phân loại hai loại hoa trong tập dữ liệu Iris: Iris-versicolor và Iris-virginica.

2. Chuẩn bị Dữ Liệu

2.1. Dữ liệu

Chúng ta sử dụng tập dữ liệu Iris từ file "iris.csv" để thực hiện bài toán phân loại. Dữ liệu này bao gồm các đặc trưng của ba loài hoa Iris khác nhau.

2.2. Tiền xử lý dữ liệu

Đổi tên cột thành 'x1', 'x2', 'x3', 'x4', và 'y' để thuận tiện trong việc làm việc với dữ liệu.

Chọn chỉ mục từ 50 đến 150 để tạo một tập dữ liệu con chứa hai loại hoa: Iris-versicolor và Iris-virginica.

Gán nhãn 0 cho Iris-versicolor và 1 cho Iris-virginica.

Chia dữ liệu thành tập huấn luyện và tập kiểm tra với tỷ lệ 70% - 30%.

3. Hiển thị Dữ Liệu

Chúng ta trực quan hóa dữ liệu bằng cách sử dụng biểu đồ scatter plot. Dưới đây là biểu đồ cho tập huấn luyện và tập kiểm tra:

3.1. Biểu đồ tập huấn luyện

Trên biểu đồ này, chúng ta có hai lớp: 'class 0' và 'class 1' tương ứng với Iris-versicolor và Iris-virginica.

Biểu đồ này giúp chúng ta thấy rằng dữ liệu của hai lớp này có thể được phân tách tương đối tốt bằng một đường thẳng.

3.2. Biểu đồ tập kiểm tra

Trên biểu đồ này, chúng ta cũng có hai lớp: 'class 0' và 'class 1'.

Biểu đồ này thể hiện tập dữ liệu kiểm tra và sẽ được sử dụng để kiểm tra hiệu suất của mô hình Perceptron sau khi được huấn luyện.

4. Thuật Toán Perceptron

4.1. Cài đặt thuật toán

Chúng ta đã cài đặt thuật toán Perceptron bằng Python. Đây là một số điểm chính của cài đặt:

Tạo một class Perceptron với các phương thức \_\_init\_\_, output, predict, và fit.

Trong phương thức fit, chúng ta thực hiện việc huấn luyện mô hình Perceptron trên dữ liệu huấn luyện.

Sử dụng learning rate (lr) và số lần lặp (n\_iter) để điều chỉnh quá trình huấn luyện.

4.2. Huấn luyện mô hình

Chúng ta đã huấn luyện mô hình Perceptron trên tập dữ liệu huấn luyện. Dự đoán được thực hiện bằng cách sử dụng phương thức predict. Các kết quả sau khi huấn luyện được hiển thị trên biểu đồ dưới đây:

Trong biểu đồ này, trục x là số lần lặp và trục y là số lần lỗi.

Chúng ta thấy rằng sau một số lần lặp, thuật toán học được một đường phân chia tốt giữa hai lớp dữ liệu.

5. Kết Quả và Đánh Giá

Chúng ta sử dụng mô hình Perceptron đã được huấn luyện để dự đoán các mẫu trong tập kiểm tra và đánh giá hiệu suất của mô hình. Kết quả được trực quan hóa trên biểu đồ sau:

Trong biểu đồ này, các điểm dữ liệu được dự đoán thuộc lớp 0 hoặc 1 và được so sánh với lớp thực tế trong tập kiểm tra.

6. Kết Luận

Trong báo cáo này, chúng tôi đã thực hiện việc huấn luyện một mô hình Perceptron trên tập dữ liệu Iris để thực hiện bài toán phân loại hai loại hoa. Chúng tôi đã thực hiện tiền xử lý dữ liệu, hiển thị dữ liệu, cài đặt thuật toán Perceptron, và đánh giá hiệu suất của mô hình.

Kết quả cho thấy rằng mô hình Perceptron có thể hoạt động tốt trên tập dữ liệu này và có khả năng phân chia hai lớp dữ liệu một cách tương đối tốt. Tuy nhiên, mô hình này là một ví dụ cơ bản và có thể cần được cải thiện để xử lý các tập dữ liệu phức tạp hơn.