**I. MLP:**

**I.1: Định nghĩa:**

MLP (Multilayer Perceptron) là một kiến trúc mạng nơ-ron nhân tạo đơn giản nhất và được sử dụng phổ biến trong các bài toán học máy.

MLP là một mạng nơ-ron truyền thẳng (feedforward neural network), có khả năng học các mối quan hệ phức tạp giữa các đặc trưng đầu vào và đầu ra.

Kiến trúc MLP bao gồm một hoặc nhiều lớp ẩn (hidden layer), mỗi lớp ẩn có nhiều nơ-ron (neuron) kết nối với các nơ-ron của lớp trước và lớp sau.

**I.2. Cấu trúc MLP:**

\* Cấu trúc của MLP (Multilayer Perceptron) gồm :

- Lớp đầu vào (input layer):

+ Chứa các nơ-ron đại diện cho các đặc trưng của dữ liệu đầu vào.

- Các lớp ẩn (hidden layers):

+ Là các lớp nơ-ron nằm giữa lớp đầu vào và lớp đầu ra.

+ Mỗi lớp ẩn chứa nhiều nơ-ron, và mỗi nơ-ron trong lớp ẩn kết nối với tất cả các nơ-ron trong lớp trước đó và lớp sau đó.

- Lớp đầu ra (output layer):

+ Chứa các nơ-ron đại diện cho đầu ra của mô hình

- Trọng số (weights):

+ Là các giá trị được gán cho các kết nối giữa các nơ-ron trong mạng.

- Ngưỡng (bias):

+ Là các giá trị được thêm vào trước khi áp dụng hàm kích hoạt (activation function) cho mỗi nơ-ron. Ngưỡng giúp điều chỉnh giá trị đầu ra của mỗi nơ-ron để đưa ra dự đoán chính xác hơn.

- Hàm kích hoạt (activation function):

+ Là một hàm phi tuyến được áp dụng lên tổng trọng số đầu vào của mỗi nơ-ron để tính toán đầu ra của nơ-ron đó.

- Hàm mất mát (loss function):

+ Là một hàm đo lường sai số giữa kết quả dự đoán và kết quả đích mong muốn.

**I.3. Giải thuật toán sử dụng MLP:**

**B1**:Chuẩn bị dữ liệu:

- Tiền xử lý dữ liệu: chuẩn hóa, chuyển đổi định dạng dữ liệu, xử lý dữ liệu thiếu hoặc nhiễu.

- Chia tập dữ liệu thành tập huấn luyện và tập kiểm tra (hoặc tập cross-validation).

**B2**: Xác định cấu trúc mạng nơ-ron:

- Xác định số lượng lớp ẩn, số lượng nơ-ron trong mỗi lớp ẩn.

- Chọn hàm kích hoạt cho các nơ-ron trong mạng.

- Chọn hàm mất mát để đánh giá hiệu quả của mô hình.

- Chọn phương pháp tối ưu hóa để cập nhật trọng số trong mạng.

**B3**: Huấn luyện mô hình:

- Khởi tạo các trọng số trong mạng nơ-ron.

- Truyền dữ liệu qua mạng nơ-ron, tính toán đầu ra và đánh giá hiệu quả của mô hình bằng hàm mất mát.

- Tính toán đạo hàm của hàm mất mát theo các trọng số trong mạng nơ-ron.

- Cập nhật các trọng số trong mạng nơ-ron bằng phương pháp tối ưu hóa.

- Lặp lại quá trình truyền ngược và cập nhật trọng số cho đến khi đạt được điều kiện dừng.

**B4**: Đánh giá mô hình:

- Đánh giá hiệu quả của mô hình trên tập kiểm tra hoặc tập cross-validation.

- Đánh giá các chỉ số đánh giá như độ chính xác, độ phân loại, độ lỗi, độ đo F1, AUC.

**B5**: Sử dụng mô hình:

- Sử dụng mô hình để dự đoán trên dữ liệu mới.

- Có thể điều chỉnh các tham số của mô hình để cải thiện hiệu quả dự đoán.

**I.4. Ưu và Nhược điểm MLP:**

**\* Ưu:**

- Khả năng học các mô hình phức tạp

- Khả năng học phi tuyến

- Có thể sử dụng cho nhiều loại bài toán

- Có tính linh hoạt

**\* Nhược:**

- Dễ bị overfitting

- Cần nhiều dữ liệu

- Tốc độ huấn luyện chậm

**I.5. Ứng dụng thực tế:**

- Nhận dạng khuôn mặt và giọng nói.

- Dự đoán giá cổ phiếu và các chỉ số tài chính.

- Xác định các thuộc tính của khách hàng và dự đoán hành vi mua hàng.

- Phân loại hình ảnh và video.

- Tự động điều khiển robot và xe tự lái.

**II.Thuật toán Regression:**

**II.1. Định nghĩa:**

Thuật toán Regression là một lớp các thuật toán trong Machine Learning được sử dụng để giải quyết bài toán dự đoán giá trị của một biến liên tục dựa trên một hoặc nhiều biến đầu vào.

Mục tiêu của thuật toán Regression là tìm ra một mô hình toán học (hàm số) phù hợp nhất để mô tả mối quan hệ giữa các biến đầu vào và biến đầu ra trong tập dữ liệu cho trước. Mô hình được tạo ra từ quá trình học máy (machine learning) dựa trên các phương pháp và công cụ phân tích dữ liệu để tìm ra các tham số tối ưu của mô hình.

**II.2. Các thuật toán phổ biến trong Regression:**

\* Linear Regression:

- để tìm mối quan hệ tuyến tính giữa biến đầu vào và biến đầu ra.

\* Polynomial Regression:

- thể hiện mô hình hóa mối quan hệ giữa biến đầu vào và biến đầu ra.

\* Ridge Regression:

- để giảm overfitting bằng cách giới hạn giá trị của các hệ số

\* Lasso Regression:

- cũng để giảm overfitting nhưng dùng hàm mất mát khác, giúp giảm số lượng hệ số không cần thiết trong mô hình.

\* Decision Tree Regression:

- để mô hình hóa mối quan hệ giữa biến đầu vào và biến đầu ra

\* Random Forest Regression:

- để tăng tính ổn định và độ chính xác của mô hình.

\* Support Vector Regression (SVR):

- để tìm một đường cong tốt nhất để mô hình hóa mối quan hệ giữa biến đầu vào và biến đầu ra.

**II.3. Giải thuật Regression:**

1. Chuẩn bị dữ liệu:

- Bước đầu tiên là thu thập và chuẩn bị dữ liệu cho mô hình

2. Chọn phương pháp Regression:

\* Gồm:

- Linear Regression, Polynomial Regression, Ridge Regression, Lasso Regression, Elastic Net Regression,...

3. Chia dữ liệu thành tập huấn luyện và tập kiểm tra:

- Tập dữ liệu sẽ được chia thành hai phần:

+ Huấn luyện mô hình

+ Kiểm tra độ chính xác của mô hình

4. Huấn luyện mô hình:

- Tìm kiếm các siêu tham số tối ưu cho phương pháp Regression đã chọn.

5. Đánh giá mô hình:

- Sử dụng tập dữ liệu kiểm tra để đánh giá độ chính xác của mô hình

- Các phương pháp đánh giá mô hình bao gồm độ chính xác, sai số trung bình bình phương (MSE), sai số trung bình tuyệt đối (MAE)

6. Tinh chỉnh mô hình:

- Thay đổi các siêu tham số hoặc sử dụng phương pháp Regularization.

7. Sử dụng mô hình:

- Sử dụng nó để dự đoán giá trị của biến đầu ra cho các giá trị của biến đầu vào mới.

**II.4. Ưu và Nhược điểm của thuật toán Regression**

**\* Ưu điểm:**

- Đơn giản và dễ hiểu

- Tính linh hoạt trong xử lí các bài toán khác nhau

- Tính ổn định( đưa ra dữ liệu dự đoán ổn định)

**\* Nhược điểm:**

- Dễ bị ảnh hưởng bởi nhiễu, Dễ bị overfitting

- Không thể xử lý được các biến không liên tục

- Giới hạn trong mô hình hóa các mối quan hệ không tuyến tính

**II.5. Ứng dụng thực tế:**

- Giá cổ phiếu

- Giá nhà đất

- Dự báo doanh số bán hàng

- Dự báo thời tiết

- Dự báo tình trạng sức khỏe