Neural Network

Họ và tên: Mai Minh Quân

MSSV: 20225661

Source:

* Machine Learning cơ bản – Vũ Hữu Tiệp
* Deep Learning cơ bản – Nguyễn Thanh Tuấn

I. Giới thiệu về Neural Network

Neural Network (NN) là một mô hình tính toán lấy cảm hứng từ cách bộ não con người hoạt động. NN bao gồm nhiều neuron nhân tạo kết nối với nhau theo một cấu trúc phân tầng và được sử dụng để nhận diện mẫu, phân loại dữ liệu, dự đoán xu hướng, và nhiều ứng dụng khác.

Mạng nơ-ron có khả năng học từ dữ liệu bằng cách điều chỉnh các trọng số và bias của các kết nối giữa các neuron để tối ưu hóa đầu ra của mô hình.

# II. Cấu trúc của Neural Network

1. Mô hình tổng quát

A diagram of a network

AI-generated content may be incorrect.

* Layer đầu tiên là input layer, các layer ở giữa được gọi là hidden layer, layer cuối cùng được gọi là output layer. Các hình tròn được gọi là node.
* Mỗi mô hình luôn có 1 input layer, 1 output layer, có thể có hoặc không các hidden layer. Tổng số layer trong mô hình được quy ước là số layer - 1 (không tính input layer).
* Ví dụ như ở hình trên có 1 input layer, 2 hidden layer và 1 output layer. Số lượng layer của mô hình là 3 layer.
* Mỗi node trong hidden layer và output layer:
* Liên kết với tất cả các node ở layer trước đó với các hệ số riêng.
* Mỗi node có 1 hệ số bias riêng.
* Diễn ra 2 bước: tính tổng linear và áp dụng activation function.

1. Công thức toán học

Mỗi nơ-ron thực hiện hai bước chính:

1. Tính tổng có trọng số của đầu vào

Trong đó:

* là ma trận trọng số giữa lớp và lớp .
* là đầu ra của lớp trước
* là bias của lớp hiện tại

- Bias

* Bias là một hệ số trong Neural Network, thường được ký hiệu là b, có vai trò tương tự như hệ số tung độ gốc trong phương trình đường thẳng.
* Bias cho phép đường thẳng (hoặc siêu phẳng trong không gian nhiều chiều) dịch chuyển, giúp mô hình biểu diễn các mối quan hệ phức tạp hơn trong dữ liệu. Nếu không có bias, đường thẳng sẽ luôn đi qua gốc tọa độ, làm giảm khả năng tổng quát hóa của mô hình.

- Hệ số Weight

* Weight, thường ký hiệu là w hoặc W, là các hệ số nhân với các giá trị đầu vào để xác định mức độ ảnh hưởng của từng đầu vào đến đầu ra của một nơ-ron.
* Mỗi liên kết giữa các nơ-ron trong mạng Neural Network có một weight tương ứng. Các weights thường được tổ chức thành các ma trận.

1. Áp dụng hàm kích hoạt

Trong đó là activation function, giúp mô hình học các đặc trưng phi tuyến

1. Activation Function
2. Tổng quan về Activation Function

Activation function là một hàm toán học được áp dụng lên đầu ra của một nơ-ron sau khi đã tính tổng tuyến tính (tổng của các đầu vào nhân với weights và cộng với bias).

Activation function tạo ra tính phi tuyến cho mô hình Neural Network, cho phép nó học được các mối quan hệ phức tạp trong dữ liệu. Nếu không có activation function phi tuyến, mạng Neural Network sẽ chỉ có thể biểu diễn các hàm tuyến tính, bất kể có bao nhiêu lớp.

1. Các loại Activation Function phổ biến

**-** Hàm Sigmoid: (đầu ra từ 0 đến 1)

**A graph of a function

AI-generated content may be incorrect.**

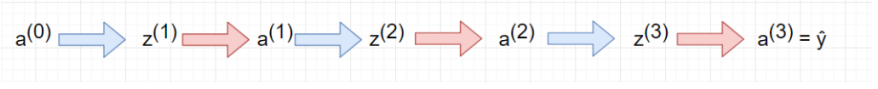
* Hàm Tanh: (đầu ra từ -1 đến 1, có tính zero-centered)

A graph of a function

AI-generated content may be incorrect.

* Hàm ReLU (Rectified Linear Unit): (đơn giản, tính toán nhanh, giúp huấn luyện mạng sâu hiệu quả)

1. Quá trình Feedforward



Feedforward là quá trình truyền dữ liệu từ đầu vào đến đầu ra thông qua các lớp nơ-ron.

* Nhận dữ liệu đầu vào từ input layer.
* Tính toán từng tầng theo công thức:
* Xuất kết quả dự đoán từ output layer.

1. Thuật toán Backpropagation

A math equations on a grid

AI-generated content may be incorrect.

- Backpropagation là thuật toán giúp tối ưu trọng số bằng cách cập nhật dựa trên đạo hàm của Loss Function.

- Backpropagation là thuật toán lan truyền ngược để cập nhật trọng số bằng cách tính toán gradient của Loss Function với từng trọng số.

- Các bước trong Backpropagation

* Feedforward: Tính đầu ra của mỗi nơ-ron.
* Tính lỗi ở đầu ra: So sánh với nhãn thực tế bằng Loss Function.
* Tính Gradient của Loss Function: Dùng đạo hàm chuỗi để lan truyền lỗi ngược về các tầng trước.
* Cập nhật trọng số bằng Gradient Descent.