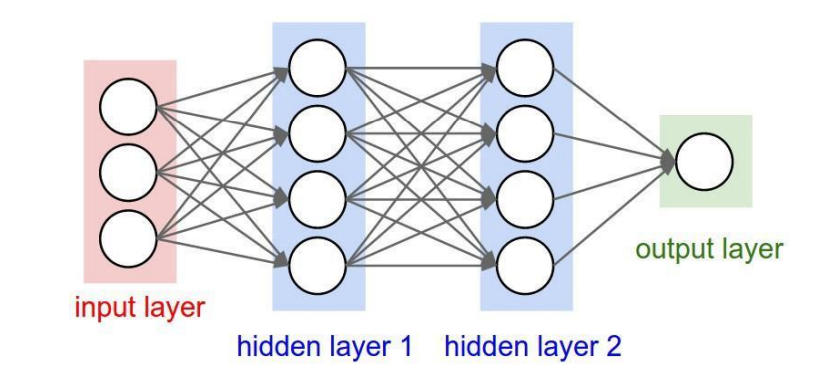
**TÓM TẮT VỀ NEURAL NETWORK VÀ TRAINING NEURAL NETWORK**

* Tổng quan về cấu trúc mô hình Neural network: Gồm 3 phần lớp chính:



* Input layer: Nhận các dữ liệu đầu vào
* Hidden layers: Thực hiện các phép tính toán và xử lý dữ liệu cho từ đầu vào để truyền kết quả đến lớp phía sau.
* Output layer: Trả về các kết quả đầu ra mong muốn.
* Trong các hidden layers, ta cần quan tâm đến trọng số liên kết giữa các nơron và hàm kích hoạt tại các nơ ron đó.

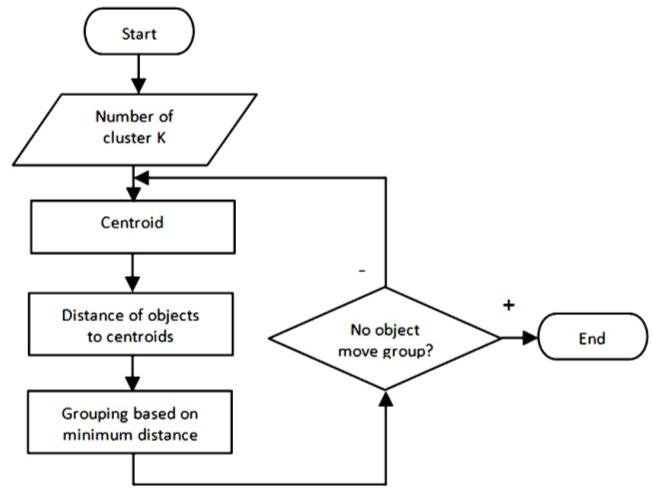
A diagram of a diagram

Description automatically generated

* Mỗi một liên kết giữa các nơron được gắn liền với 1 trọng số.
* Mỗi 1 nơron sẽ có 1 hàm kích hoạt activation function. Kết quả hàm đó sẽ được truyền đi sang lớp kế tiếp.
* Với mỗi 1 node trong hidden layer, ta thực hiện lần lượt 2 bước : Tính tổng tuyến tính tổ hợp các node trong layer trước nhân với trọng số của các liên kết tương ứng rồi cộng với hằng số bias. Kết quả đó được đưa vào hàm kích hoạt để cho ra kết quả của node đó.
* Dữ liệu từ input layer cứ như vậy sẽ được tính feedforward qua các hidden layer để ra được output layer.
* Sau khi có giá trị dữ đoán thì mình sẽ tạo loss function và dùng thuật toán backpropagation để tính đạo hàm của loss với các tham số W, b, rồi dùng thuật toán gradient descent để cập nhật hệ số và tối ưu loss function.

**TÓM TẮT VỀ K-MEANS CLUSTERING VÀ APPLICATIONS**

* K-means clustering (phân cụm K-means) là thuật toán unsupervised learning dùng để nhóm cụm các điểm dữ liệu có tính chất giống nhau.
* Hoạt động của thuật toán:
* Chọn số nhóm K cần chia, lựa chọn này quan trọng ảnh hương đến đầu ra kết quả
* Random vị trí trung tâm (centroid) của các cụm Ki trên hệ tọa độ
* Với mỗi điểm dữ liệu có sẵn, tính toán khoảng cách tới từng điểm trung tâm của các cum Ki đó, sau đó gán nhãn điểm đó gắn với cụm mà có khoảng cách gần nhất.
* Tính toán lại trung tâm của từng cụm bằng cách lấy trung bình cộng của tất cả các điểm dữ liệu được gán nhãn gắn liền với cụm đó. Giá trị trung bình mới trở thành tọa độ trung tâm mới của cụm.
* Lặp lại các bước trên đến khi vị trí trung tâm không còn di chuyển quá đáng kể thì ta dừng lại.



* Một số điểm lưu ý về thuật toán:
* Thuật toán không giám sát 🡪 Dữ liệu không cần dán nhãn sẵn, mà nó sẽ tự tìm khuôn mẫu trong dữ liệu
* Việc lựa chọn vị trí centroids ban đầu có thể ảnh hưởng đến kết quả chia cụm lúc cuối cùng.
* Thuật toán nhạy cảm với các dữ liệu ngoài phạm vi và dữ liệu nhiễu 🡪 cần tiền xử lý loại bỏ để tránh ảnh hưởng độ chính xác của thuật toán.