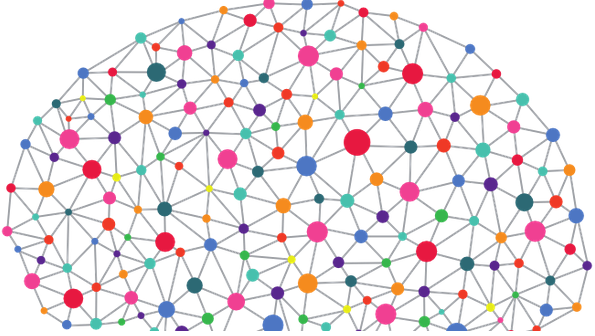
**Ứng dụng thuật toán máy   
học phân loại phương tiện giao thông**

1. **Các nội dung và thuật toán nhóm dự định tìm hiểu.**
2. ***Tìm hiểu về deep learning.***
3. ***TÌm hiểu về Neural Networks.***
4. ***Tìm hiểu thuật toán CNN.***
5. ***Áp dụng thuật toán CNN vào bài toán phân loại phương tiện giao thông.***
6. **Deep Learning là gì?**

****

* Deep Learning là một thuật toán dựa trên một số ý tưởng từ não bộ tới việc tiếp thu nhiều tầng biểu đạt, cả cụ thể lẫn trừu tượng, qua đó làm rõ nghĩa của các loại dữ liệu. Deep Learning được ứng dụng trong nhận diện hình ảnh, nhận diện giọng nói, xử lý ngôn ngữ tự nhiên
* Hiện nay rất nhiều các bài toán nhận dạng sử dụng deep learning để giải quyết do deep learning có thể giải quyết các bài toán với số lượng lớn, kích thước đầu vào lớn với hiệu năng cũng như độ chính xác vượt trội so với các phương pháp phân lớp truyền thống
* Những năm gần đây, ta đã chứng kiến được nhiều thành tựu vượt bậc trong ngành Thị giác máy tính (Computer Vision). Các hệ thống xử lý ảnh lớn như Facebook, Google hay Amazon đã đưa vào sản phẩm của mình những chức năng thông minh như nhận diện khuôn mặt người dùng, phát triển xe hơi tự lái hay drone giao hàng tự động.
* Convolutional Neural Network (CNNs – Mạng nơ-ron tích chập) là một trong những mô hình Deep Learning tiên tiến giúp cho chúng ta xây dựng được những hệ thống thông minh với độ chính xác cao như hiện nay

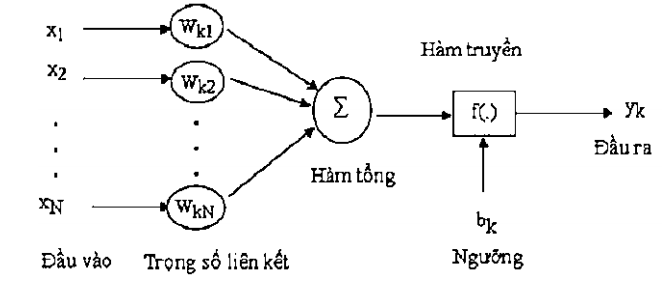
**II. TÌm hiểu về Neural Networks.**

**Tổng quan:**

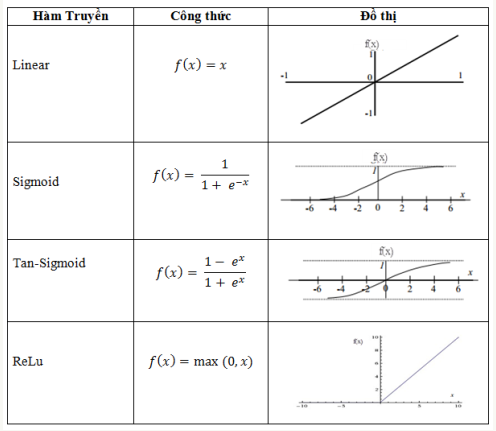
Mạng neural nhân tạo (Artificial Neural Network : ANN), gọi tắt neural network là mô hình xử lý thông tin mô phỏng hoạt động của các hệ neuron sinh học mà cụ thể hơn ở đây là bộ não con người. Trong đó, thành phần cơ bản của ANN là neural nhân tạo có cách thức hoạt động và xử lý tương tự neuron sinh học. ANN được hình thành từ số lượng lớn các neural được liên kết với nhau theo cấu trúc từng tầng (layer), các neural kết nối với nhau giữa các tầng thông qua trọng số liên kết (weight).

1. Neural nhân tạo

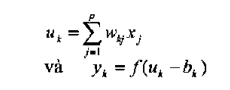
Cấu trúc:



* Các thành phần của mạng neural nhân tạo bao gồm:
* Tập các đầu vào: Là các tín hiệu vào (Input signals) của neural, các tính hiệu này thường được đưa vào dưới dạng một vector N chiều.
* Tập các liên kết: Mỗi liên kết được thể hiện bởi một trọng số liên kết - Synaptic weight. Trọng số liên kết giữa các tính hiệu vào thứ j với neural k thường được kí hiệu là . Thông thường, các trọng số này được khởi tạo một cách ngẫu nhiên ở thời điểm khởi tạo mạng và được cập nhật liên tục trong quá trình học mạng.
* Hàm tổng: Thường dùng để tính tổng của tích các đầu vào với trọng số liên kết của nó.
* Ngưỡng: Ngưỡng này thường được đưa vào như một thành phần của hàm truyền.
* Hàm truyền: Hàm này được dùng để giới hạn phạm vi đầu ra của mỗi neural. Nó nhận đầu vào là kết quả của hàm tổng và ngưỡng.Thông thường, phạm vi đầu ra của mỗi neural được giới hạn trong đoạn [0,1] hoặc [-1, 1]. Các hàm truyền rất đa dạng, có thể là các hàm tuyến tính hoặc phi tuyến. Việc lựa chọn hàm truyền nào là tuỳ thuộc vào từng bài toán và kinh nghiệm của người thiết kế mạng. Một số hàm truyền thường sử dụng trong các mô hình mạng nơron được đưa ra trong bảng sau:



* Đầu ra: Là tín hiệu đầu ra của một neural, với mỗi neural sẽ có tối đa một đầu ra.
* Xét về mặt toán học, cấu trúc của một neural k, được mô tả bằng cặp biểu thức:



Trong đó:

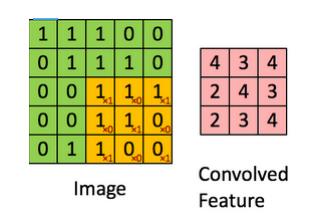
* **,,,...,**: là các tín hiệu vào.
* **,,…,** là các trọng số liên kết của neural thứ k.
* là hàm tổng.
* là một ngưỡng.
* f là hàm truyền.
* là tín hiệu đầu ra của neural.

**=>** Như vậy neural nhân tạo nhận các tín hiệu đầu vào, xử lý (nhân các tín hiệu này với trọng số liên kết, tính tổng các tích thu được rồi gửi kết quả tới hàm truyền), và cho một tín hiệu đầu ra ( là kết quả của hàm truyền).

**III. Tìm hiểu về Convolutional trong CNN.**

1. **Convolutional là gì?**

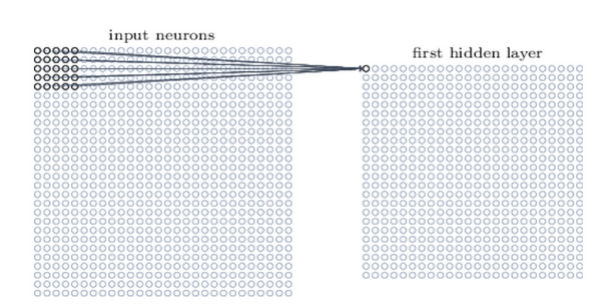
* Là một cửa sổ trượt (Sliding Windows) trên một ma trận như mô tả hình dưới:



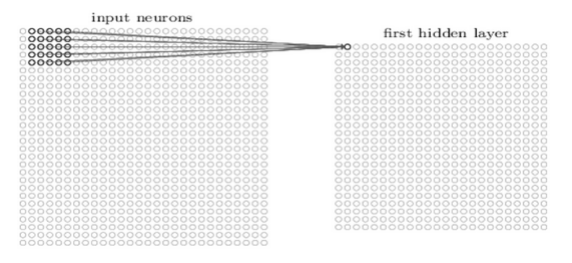
* Trong hình ảnh ví dụ trên, ma trận bên trái là một hình ảnh trắng đen được số hóa.
* Ma trận convolved feature là kết quả nhân từng ma trận có kích thước 3\*3 trong ma trận lớn 5\*5 với một ma trận 3\*3 cho trước.
* Ma trận có kích thước 3\*3 cho trước được gọi là sliding windows hay còn được gọi là kernel, filter hoặc feature detect.

1. **Cấu trúc mạng CNN.**

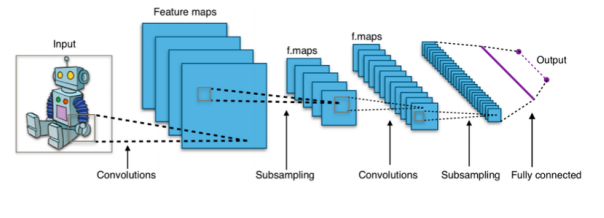
* Mạng CNN là một tập hợp các lớp Convolution chồng lên nhau và sử dụng các hàm nonlinear activation.
* Mỗi một lớp sau khi thông qua các hàm kích hoạt sẽ tạo ra các thông tin trừu tượng hơn cho các lớp tiếp theo.
* Các layer liên kết được với nhau thông qua cơ chế convolution.
* Layer tiếp theo là kết quả convolution từ layer trước đó, nhờ vậy mà ta có được các kết nối cục bộ. Như vậy mỗi neuron ở lớp kế tiếp sinh ra từ kết quả của filter áp đặt lên một vùng ảnh cục bộ của neuron trước đó.
* Mỗi một lớp được sử dụng các filter khác nhau.
* Trong quá trình huấn luyện mạng (traning) CNN tự động học các giá trị qua các lớp filter dựa vào cách thức mà bạn thực hiện. Ví dụ trong tác vụ phân lớp ảnh, CNNs sẽ cố gắng tìm ra thông số tối ưu cho các filter tương ứng theo thứ tự raw pixel > edges > shapes > facial > high-level features. Layer cuối cùng được dùng để phân lớp ảnh.
* Trong mô hình CNN có 2 khía cạnh cần quan tâm là tính bất biến (Location Invariance) và tính kết hợp (Compositionality).
* Mạng CNN sử dụng 3 ý tưởng cơ bản:
* các trường tiếp nhận cục bộ (local receptive field).
* Đầu vào của mạng CNN là một ảnh. Ví dụ như ảnh có kích thước 28x28 thì tương ứng đầu vào là một ma trận có 28x28 và giá trị mỗi điểm ảnh là một ô trong ma trận.
* các bước tạo ra 1 hidden layer bằng các cách sau:
* Tạo ra neuron ẩn đầu tiên trong lớp ẩn 1.



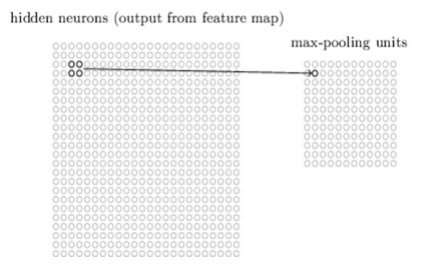
* Dịch filter qua bên phải một cột sẽ tạo được neuron ẩn thứ 2.



* Đối với bài toán nhận dạng ảnh người ta thường gọi ma trận lớp đầu vào là feature map, trong nhận dạng ảnh chúng ta cần nhiều hơn một feature map.
* trọng số xác định các đặc trưng là shared weight và độ lệch xác định một feature map là shared bias.
* trọng số chia sẻ (shared weights)
* Đầu tiên, các trọng số cho mỗi filter (kernel) phải giống nhau.
* việc map từ input layer sang hidden layer là một feature map.
* Chúng ta thấy mỗi fearture map cần 25 = 5x5 shared weight và 1 shared bias. Như vậy mỗi feature map cần 5x5+1 = 26 tham số. Như vậy nếu có 10 feature map thì có 10x26 = 260 tham số.
* trong các mô hình, mô hình có số lượng tham số ít hơn thì nó sẽ chạy nhanh hơn.
* Tóm lại, một convolutional layer bao gồm các feature map khác nhau. Lợi ích lớn nhất của trọng số chia sẻ là giảm tối đa số lượng tham số trong mạng CNN.
* tổng hợp (pooling).
* Lớp pooling thường được sử dụng ngay sau lớp convulational để đơn giản hóa thông tin đầu ra để giảm bớt số lượng neuron.



* Thủ tục pooling phổ biến là max-pooling, thủ tục này chọn giá trị lớn nhất trong vùng đầu vào 2x2.



* Như vậy qua lớp Max Pooling thì số lượng neuron giảm đi phân nửa. Trong một mạng CNN có nhiều Feature Map nên mỗi Feature Map chúng ta sẽ cho mỗi Max Pooling khác nhau.
* Cuối cùng ta đặt tất cả các lớp lại với nhau thành một CNN với đầu ra gồm các neuron với số lượng tùy bài toán.