

**TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP. HỒ CHÍ MINH**  
**KHOA ĐÀO TẠO CHẤT LƯỢNG CAO**  
**NGÀNH CÔNG NGHỆ THÔNG TIN**



# **MACHINE LEARNING**

**Đề tài: Handwritten Digit Recognition using Convolutional Neural Networks**

**GVHD : Vũ Quang Huy**

**SVTH :**

**Huỳnh Kiến Minh**

**15110079**

**Tôn Nữ Minh Châu**

**15110019**

**Tp. Hồ Chí Minh, tháng năm 2018**

## Mục lục

1. Introduction – Giới thiệu .....	2
1.1. Reason – Lý do.....	2
1.1.1. Scope của đề tài : .....	2
1.1.2. Đề tài có thể áp dụng với Convolutional Neural Network .....	2
1.2. Meaning – Ý Nghĩa .....	2
2. Goal – Kết quả.....	3
3. Objectives – Các mục tiêu cần đạt được .....	3
4. Theory – Cơ sở lý thuyết .....	3
4.1. Neutral Network .....	3
4.2. Activation Function.....	5
1.2.1. Sigmoid function.....	5
1.2.2. Rectified linear unit (ReLU) .....	6
1.2.3. Softmax Function .....	7
4.3. Cost Function .....	7
4.4. Thuật toán tối ưu Gradient Descent .....	7
4.5. How neural network learn .....	9
4.6. Convolution Neural Network – CNN.....	11
4.6.1. Convolutional Layer .....	11
4.6.2. Pooling Layer .....	12
4.6.3. Fully Connected Layer.....	12
5. Diagram cho bài toán Supervised Learning ( Classification).....	13
6. Mô hình Convolution Neural Network cụ thể áp dụng cho đề bài .....	14
Tài liệu tham khảo.....	14

## 1. Introduction – Giới thiệu

- Đề án: Handwritten Digit Recognition using Convolutional Neural Networks
- Limitation:
  - Đề tài: nhận diện chữ số từ 1 đến 10.
  - Kiến thức: áp dụng những kiến thức cơ bản của Neural Networks, kiến thức của một số thuật toán, activation function, cost functions để phục vụ cho bài toán nhận dạng chữ số viết.
  - Sản phẩm: sản phẩm là chương trình được viết bằng ngôn ngữ lập trình Python. Sử dụng bộ data set MNIST để đào tạo và đánh giá.

### 1.1. Reason – Lý do

Việc lựa chọn đề tài dựa trên 2 yếu tố chính

#### 1.1.1. Scope của đề tài :

Do thời gian chỉ giới hạn trong một học kỳ và phải tiếp thu nhiều kiến thức mới nên nhóm quyết định chọn một đề tài vừa sức, phù hợp với thời gian và độ phức tạp nhưng vẫn có thể áp dụng những kiến thức cơ bản đã học và tự tìm hiểu những kiến thức mới.

Với đề tài Handwritten Digit Recognition không quá phức tạp về mặt nghiệp vụ nên nhóm sẽ có nhiều thời gian hơn để chú trọng vào lý thuyết và tìm hiểu kiến thức kỹ những kiến thức mới để áp dụng vào bài toán.

#### 1.1.2. Đề tài có thể áp dụng với Convolutional Neural Network

Convolutional Neural Network là kiến thức mà nhóm có thể áp dụng cho đề tài, đây cũng là nền tảng để tiếp cận với kiến thức sâu hơn về Machine Learning như Deep Learning nên nhóm sẵn có củng cố kiến thức và tự tìm hiểu thêm về mảng kiến thức mới này.

### 1.2. Meaning – Ý Nghĩa

- Hiểu được khái quát mối quan hệ giữa kiến thức rời rạc để hoàn thành được đề tài.
- Hiểu được bản chất của các thuật toán: ưu điểm và nhược điểm, khi nào và áp dụng như thế nào?
- Rèn luyện khả năng đọc hiểu và áp dụng những kiến thức mới vào đề tài.
- Hiểu và áp dụng Convolutional Neural Network cho bài toán phân loại đơn giản.
- Có kiến thức về Python và tìm hiểu các thư viện hỗ trợ như Keras.
- Thực hành được những kiến thức tự tìm hiểu và là cơ sở để tìm hiểu sâu về lĩnh vực Artificial Intelligence, Data Analyst.

## 2. Goal – Kết quả

Xây dựng được chương trình nhận diện được chữ số viết tay đơn giản.

## 3. Objectives – Các mục tiêu cần đạt được

- Tìm hiểu tiến trình của một bài toán Machine Learning, cụ thể là bài toán Supervised Learning (Classification)
  - Training phase
  - Predict phase
- Tìm hiểu cách mà neural network learn (training process)
  - Perceptron
  - Multiple-layer Perceptron
    - Forward Propagation
    - Back-Propagation
  - Activation Function: Sigmoid, Rectified linear unit (ReLU), Softmax
    - Cách áp dụng
    - Cách chọn activation phù hợp
  - Cost Function: sự ảnh hưởng của weight, bias trong quá trình học model.
  - Thuật toán tối ưu: Gradient Descent (Batch Gradient Descent, Stochastic Gradient Descent)
- Tìm hiểu cách xử lý dữ liệu đầu vào, cụ thể là hình ảnh (image processing) – Convolution Neural Network (CNN)
  - Convolution layer
  - Pooling Layer
  - Fully-connected Layer
- Áp dụng lý thuyết vào chương trình nhận diện chữ viết số viết tay.

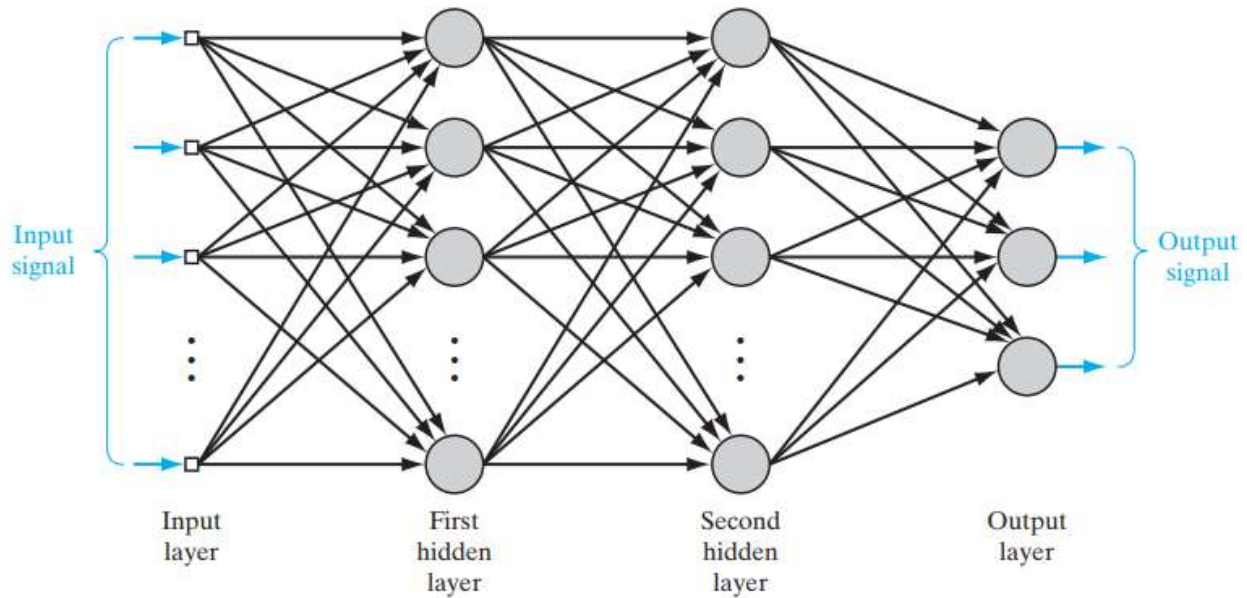
## 4. Theory – Cơ sở lý thuyết

### 4.1. Neural Network

Dựa trên mô hình liên kết giữa các tế bào thần kinh trong não bộ, mô hình neural network là tập hợp các đơn vị kết nối gọi là neuron, các neuron liên kết truyền tín hiệu cho nhau.

Neural Network được phát triển dựa trên nền tảng Perceptron Learning Algorithm (PLA) hay còn gọi là Perceptron.

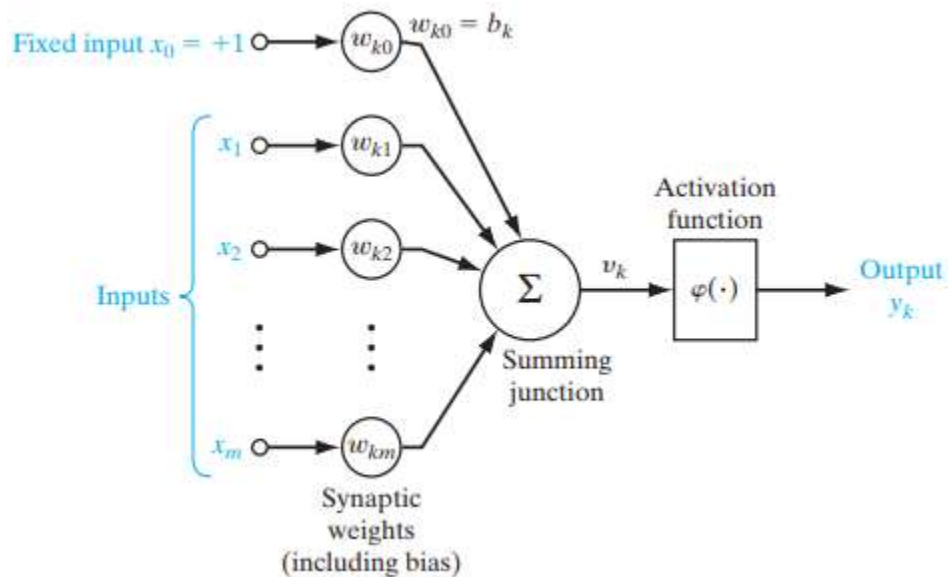
Neural network có cấu trúc Multi-layer Perceptrons. Perceptron có thể hiểu tương đương single-layer neural network.



Hình: Neural network structure

Neuron là một đơn vị xử lý thông tin cơ bản cho hoạt động của neural network

Một neuron trong bất kỳ layer nào trong network đều được kết nối với tất cả các neuron (node) trong layer trước đó – fully connected.



Hình : Mô hình non-linear của một neuron, có nhãn k

Phương trình viết lại từ mô hình:

$$v_k = \sum_{j=1}^m w_{kj} x_j$$

và

$$y_k = \varphi(v_k)$$

Với

- $x_1, x_2, \dots, x_m$  : input
- $w_{k1}, w_{k2}, \dots, w_{km}$  : trọng số tương ứng của neuron k
- $b_k$  : bias
- $\varphi(.)$  : activation function
- $y_k$  : output

#### 4.2. Activation Function

Mục đích: giới hạn phạm vi giá trị tới một giá trị hữu hạn cho phép.

Một số hàm activation thường dùng : Sigmoid, Rectified linear unit (ReLU), Softmax.

##### 1.2.1. Sigmoid function

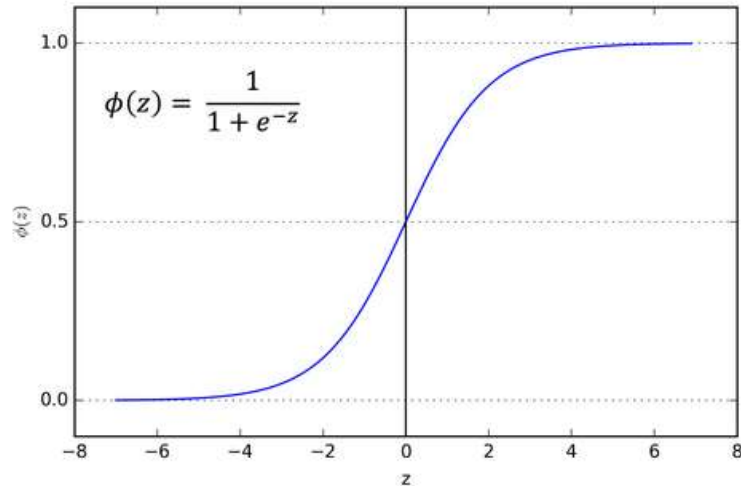
Sigmoid function là một logistic function.

Phạm vi (0, 1). Do đó, nó đặc biệt được sử dụng cho các mô hình mà phải dự đoán xác suất. Vì vậy xác suất của bất cứ điều gì chỉ tồn tại giữa phạm vi 0 và 1.

Áp dụng tốt nhất có bài toán classification với chỉ hai class.

$$f(z) = \frac{1}{1 + e^{-z}}$$

$$f'(z) = f(z)(1 - f(z))$$

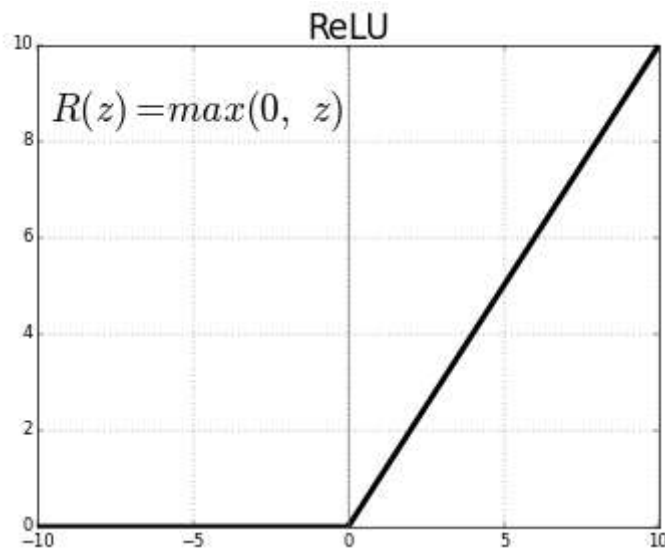


Hình: Đồ thị của Sigmoid Function

#### 1.2.2. Rectified linear unit (ReLU)

ReLU là activation function được sử dụng nhiều nhất trên thế giới hiện nay, nó được sử dụng trong Convolutional Neural Networks (CNN) hay Deep Learning.

$$f(z) = \max(0, z)$$



Hình: Đồ thị của ReLU Function

Phạm vi:  $[0, \infty)$

Với bất kỳ input có giá trị âm đưa vào cho hàm ReLU sẽ biến giá trị thành 0, do đó ảnh hưởng đến biểu đồ kết quả khi không ánh xạ các giá trị âm một cách thích hợp.

### 1.2.3. Softmax Function

Hàm softmax có output nằm trong khoảng từ 0 đến 1, giống như hàm sigmoid. Nhưng nó chia mỗi output sao cho tổng số output bằng 1.

$$f(z)_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}} \text{ for } j = 1, 2, \dots, K$$

Với  $z$ :  $K$ -dimensional vector

### 4.3. Cost Function

Cost function là độ sai lệch

$$C(w, b) = \frac{1}{2n} \sum_x \|y(x) - a\|^2$$

Với

- $n$ : số lượng training input
- $a$ : output (vector) với input  $x$
- $x$ : input (vector)

$C(w, b) \approx 0$  khi giá trị output gần như sắp xỉ bằng  $y(x)$ . Từ đó cho thấy thuật toán training hoạt động tốt hay không dựa vào việc tìm được weights and biases để  $C(w, b) \approx 0$ .

Để tránh sai số cao cần làm tối thiểu chi phí mất mát thấp nhất có thể

➔ để kiểm tra lại độ chính xác của việc học (phân loại).

Sử dụng thuật toán Gradient Descent để tối ưu hàm mất mát.

### 4.4. Thuật toán tối ưu Gradient Descent

\* Local minimum: điểm cực tiểu ( nghiệm của đạo hàm  $f'(x) = 0$ ).

\* Global minimum: điểm mà tại đó hàm số đạt cực tiểu.

Giả sử cần tìm global minimum cho hàm  $f(\theta)$ , với  $\theta$  (khởi tạo  $\theta_0 = 0$ ) là một vector.

Ta có đạo hàm của hàm mất mát tại  $\theta$  bất kỳ bằng:  $\nabla_{\theta} f(\theta)$ .

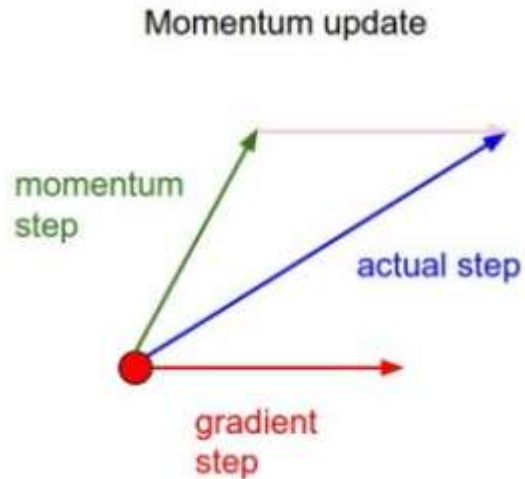
Ta có quy tắc cập nhật di chuyển:

$$\theta = \theta - \eta \nabla_{\theta} f(\theta)$$

*Các thuật toán tối ưu cho Gradient Descent*

- **Momentum**





Lượng thay đổi là tổng của hai vector: momentum vector và gradient ở thời điểm hiện tại.

Gọi lượng thay đổi ở thời điểm  $t$  là  $v_t$ , ta có vị trí mới được cập nhập:

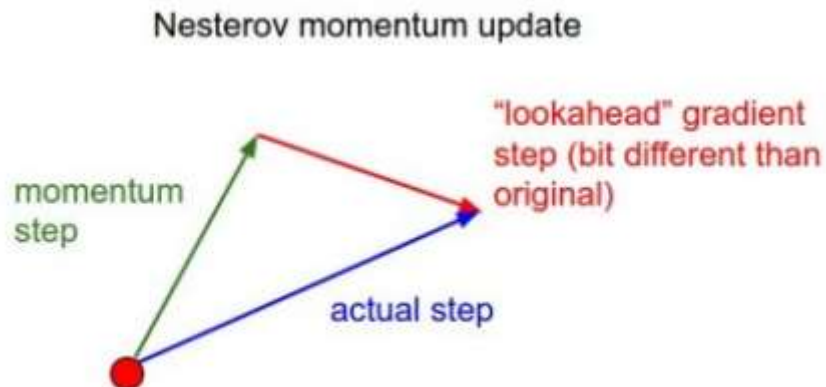
$$\theta = \theta - v_t$$

Với

$$v_t = \gamma v_{t-1} + \eta \nabla_{\theta} f(\theta)$$

Trong đó:

- $\gamma = 0.9$
  - $v_{t-1}$ : vận tốc tại thời điểm trước đó (đà).
  - $\nabla_{\theta} f(\theta)$ : là độ dốc của điểm trước đó.
- **Nesterov accelerated gradient (NAG)**



Lượng thay đổi là tổng của hai vector: momentum vector và gradient ở thời điểm được xấp xỉ là điểm tiếp theo

Công thức cập nhật của NAG như sau:

$$\theta = \theta - v_t$$

Với

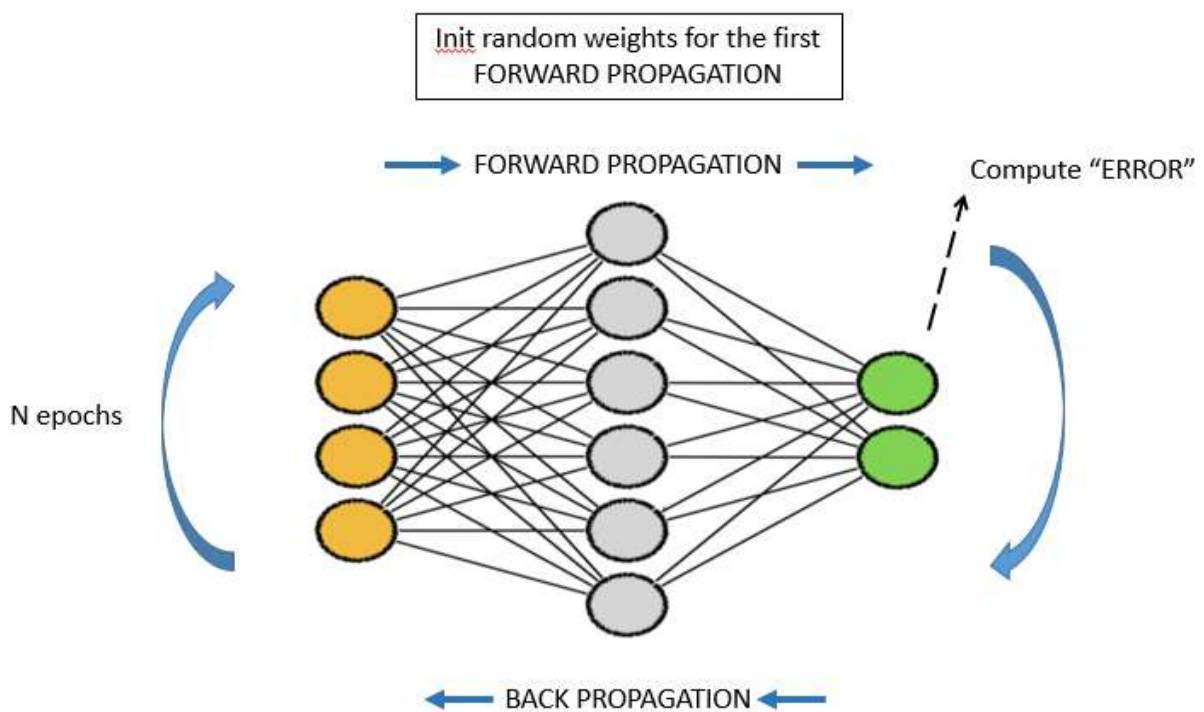
$$v_t = \gamma v_{t-1} + \eta \nabla_{\theta} f(\theta - \gamma v_{t-1})$$

## 6.2. Convolution Neural Network – CNN

Convolutional Neural Network –CNN là một dạng Artificial Neural Network, giống như Neural Network truyền thống nhận input, biến đổi input qua hidden layer nhưng khác nhau ở chỗ các layer được tổ chức theo không gian 3 chiều và những neurons trong 1 layer không liên kết đến tất cả các neuron ở layer sau

### 4.5. How neural network learn

Quá trình học của neural network có 2 phần chính là Forward Propagation và Back Propagation



Hình: Mô hình thể hiện quá trình học của neural network

\*Một Epoch là khi tất cả dữ liệu được đưa vào mạng neural network 1 lần.

Chu kỳ hoạt động của quá trình trên được cụ thể với mẫu training

$$\{(x(n), d(n))\}_{n=1}^N :$$

- 1) Initialization: khởi tạo random các trọng số ( $w$ ) trong khoảng  $(0, 1)$
- 2) Training data: đưa dữ liệu vào input layer
- 3) Forward Computation:

- Ta có input vector  $x(n)$ , desired output vector  $d(n)$
- Neuron  $j$  layer  $l$  có:

$$v_j^l(n) = \sum_i w_{ji}^l(n) \cdot y_i^{l-1}(n)$$

Với

- $y_i^{l-1}(n)$ : output của neuron  $i$  layer  $(l - 1)$  tại lần lặp thứ  $n$
- $w_{ji}^l(n)$ : trọng số của neuron  $j$  layer  $l$
- Với  $i = 0$ , ta có  $y_0^{l-1}(n) = 1$  và  $w_{j0}^l(n) = b_j^l$  (bias áp dụng cho neuron  $j$  layer  $l$ )
- Qua activation function, output của neuron  $j$  layer  $l$  bằng:

$$y_j^l(n) = \varphi_j(v_j(n))$$

- Đối với neuron  $j$  thuộc hidden layer ( $l = 1$ ) đầu tiên thì đặt:

$$y_j^0(n) = x_j(n)$$

với  $x_j(n)$  là phần tử thứ  $j$  của input vector  $x(n)$ .

- Đối với neuron  $j$  thuộc output layer ( $l = L$ ) đầu tiên thì:

$$y_j^L(n) = o_j(n)$$

- Tính độ sai lệch

$$e_j(n) = d_j(n) - o_j(n)$$

với  $d_j(n)$  là phần tử thứ  $j$  desired output vector  $d(n)$ .

- 4) Backward Computation:

Thuật toán backpropagation áp dụng để hiệu chỉnh  $\Delta w_{ji}(n)$  của trọng số  $w_{ji}(n)$ . Theo quy tắc chuỗi tính toán (chain rule), ta có đạo hàm từng phần của  $\frac{\partial E(n)}{\partial w_{ji}(n)}$  như sau:

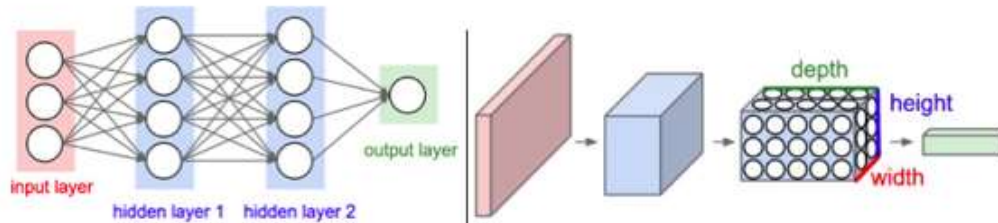
$$\frac{\partial E(n)}{\partial w_{ji}(n)} =$$

(to be continue)

- 5) Iteration: Lặp lại công việc tính toán Forward và Backward (mục 3,4 phía trên) sau mỗi epoch cho đến khi hết  $n$  epoch ( $n$  là số lần lặp được cho

trước). Bộ dữ liệu training phải trộn ngẫu nhiên sau mỗi epoch. Các thông số momentum, learning-rate đều được chỉnh khi số lần lặp tăng lên, thường là giảm.

#### 4.6. Convolution Neural Network – CNN



Hình: Mô hình Neural network (trái) và CNN (phải)

Convolutional Neural Network –CNN là một loại Artificial Neural Network. CNN dùng để xử lý các dữ liệu 2D, 3D như: image, audio, text. Cũng như Neural Network truyền thống CNN theo cấu trúc MLP gồm có: input layer, hidden layers và output layer.

Nhưng tập hidden layers trong CNN bao gồm thêm convolutional layers, pooling layers, fully connected layers.

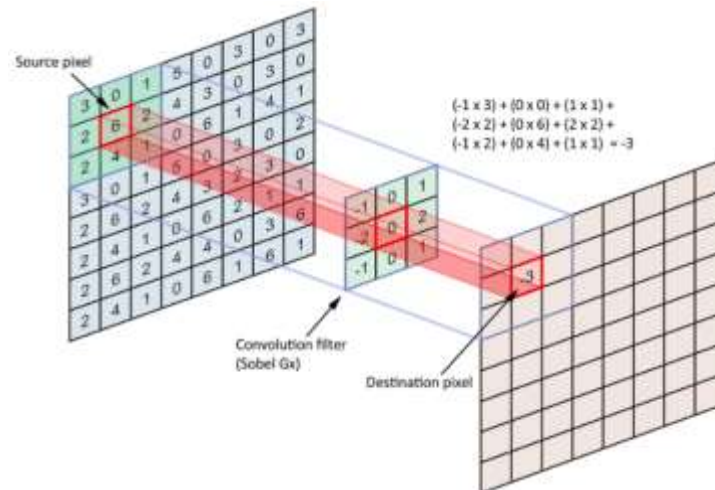
##### 4.6.1. Convolutional Layer

Mục đích: trích xuất ra các feature/đặc tính cụ thể.

Tập các feature map, mỗi feature map là 1 bản scan thu nhỏ đặc tính cụ thể của dữ liệu input.

Quá trình scan: dựa vào convolution filter/kernel, ma trận sẽ quét qua ma trận dữ liệu đầu vào, từ trái qua phải, trên xuống dưới.

Kết quả đưa qua activation function (sigmoid, relu,...), kết quả sẽ là một con số cụ thể. Tập hợp các con số này lại là 1 ma trận nữa, chính là **feature map**



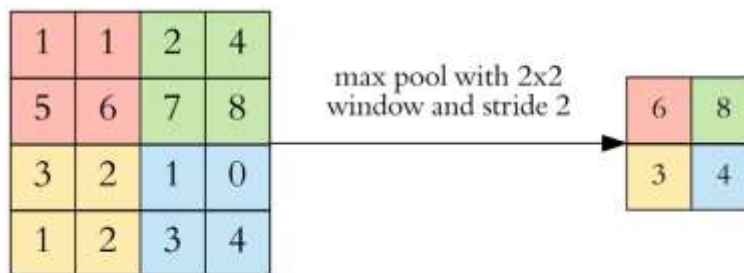
Hình: Mô tả tính toán Convolution

#### 4.6.2. Pooling Layer

Mục đích: giảm số hyperparameter cần phải tính toán, từ đó giảm thời gian tính toán, tránh overfitting.

Có 2 loại Pooling: max pooling (dùng nhiều nhất), average pooling.

Max pooling: lấy giá trị lớn nhất trong một *pooling window*.



Hình: Mô tả tính toán max pooling

#### 4.6.3. Fully Connected Layer

Mục đích: kết nối mọi neuron trong một lớp với mọi neuron trong một lớp khác.

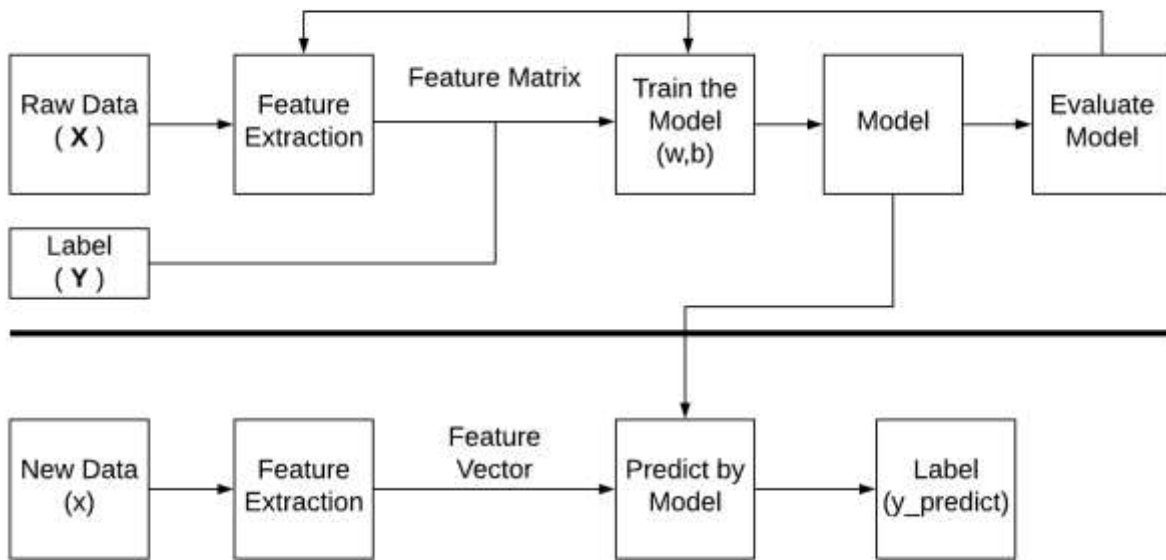
Về nguyên tắc, nó giống như multi-layer perceptron neural network truyền thống.

Thực hiện sau convolutional layers + pooling layers. Vì output của 2 layer này có kích thước 3D, còn output của fully connected layer là mảng 1D nên phải Flatten output của pooling layer cuối cùng.

Flatten chỉ đơn giản là sắp xếp khối 3D của các số thành một vector 1D ( n hàng ,1 cột).

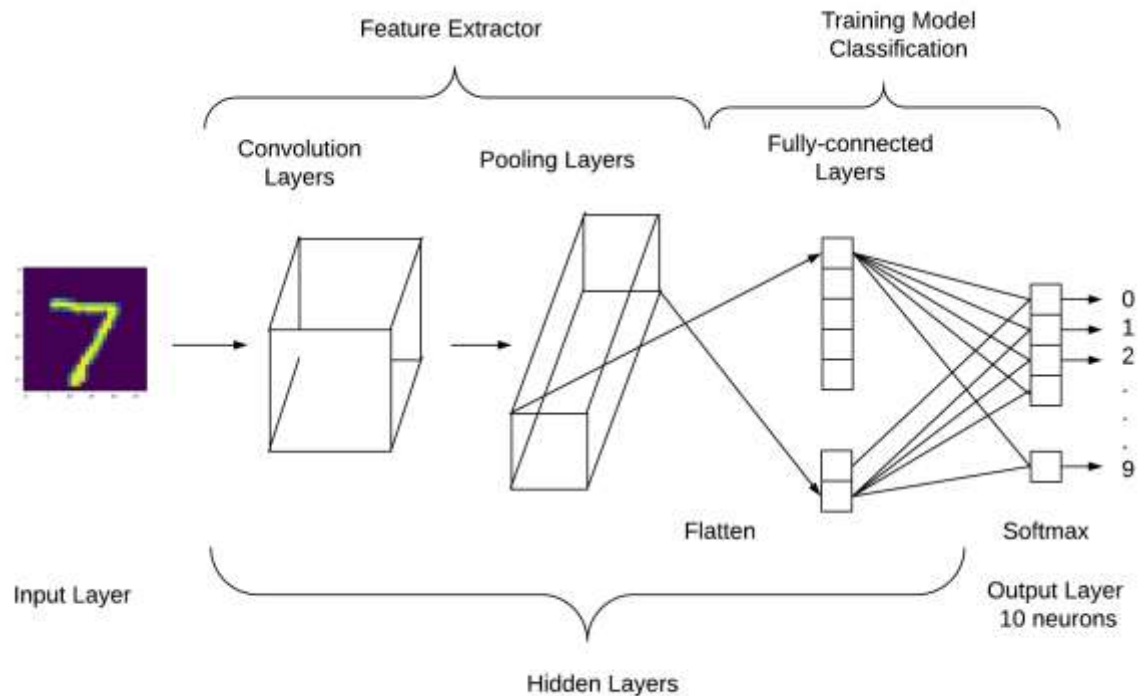
## 5. Diagram cho bài toán Supervised Learning ( Classification)

### Training Phase



### Predicting Phase

## 6. Mô hình Convolution Neural Network cụ thể áp dụng cho đề bài



- Số neuron của Output Layer bằng số class phân loại. Đề bài nhận dạng chữ số viết tay, ta có 10 số : 0, 1, 2, 3, 4, 5, 6, 7, 8, 9  
➔ Số neuron của Output Layer: 10 (neuron)

### Tài liệu tham khảo

1. Simon Haykin (2009), *Neural Networks and Learning Machines Third Edition*, Pearson
2. <https://github.com/tnmchau/TLCN2018-NeuralNetwork>
3. [https://en.wikipedia.org/wiki/Convolutional\\_neural\\_network](https://en.wikipedia.org/wiki/Convolutional_neural_network)
4. <http://cs231n.github.io/convolutional-networks/>
5. <https://towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2>