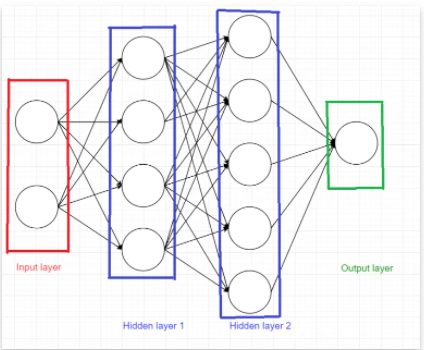
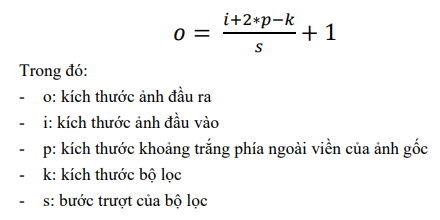
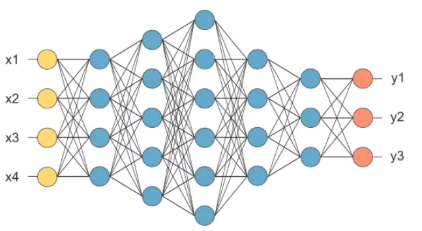
**Câu 1: Cấu trúc của một mạng CNN**

* Định nghĩa: Mạng cnn (convolutional neural network): hiểu đơn giản nó là tập hợp của các lớp Convolution chồng lên nhau và sử dụng các hàm kích hoạt phi tuyến tính (VD: Hàm ReLU, hàm tank) để kích hoạt trọng số cho các node, thông tin của node này là đầu vào của node kế tiếp (mô hình này được gọi là fully connected layer)
* Ý tưởng:
  1. Áp dụng các bộ lọc lên ảnh trước khi training mạng neural
  2. Sau khi đi qua các bộ lọc thì các đặc trưng của tấm ảnh sẽ trở lên nổi bật hơn
  3. Bộ lọc ở đây thực chất là tổ hợp của các phép nhân tích chập
  4. Kết hợp các bộ lọc lại với nhau bằng lớp tổng hợp pooling
  5. Lớp pooling sẽ gom một tập các điểm ảnh lại với nhau và cho ra một tập nhỏ hơn
* Cấu trúc của CNN:
  1. Lớp tích chập (Convolutional): Có nhiệm vụ trích xuất các đặc trưng của ảnh, output của convolutional layer đầu tiên là input của convolutional layer tiếp theo, convolutional layer bao gồm: input(dữ liệu đầu vào) và bộ lọc tích chập:
     + Bộ lọc: ví dụ một bộ lọc kích thước 3x3, di chuyển lần lượt qua từng vùng ảnh cho đến khi quét xong toàn bộ ảnh, tạo ra ảnh có kích thước nhỏ hơn hoặc bằng ảnh ban đầu. Kích thước này tùy theo khoảng trắng đc thêm ở viền
     + Tổng quan: Gỉa sử input của một convolutional layer là tensor kích thước H \* W \* D, Kernel kích thước F \* F \* D (kernel luôn có depth bằng depth của input và F là số lẻ), stride: S, padding: P.
     + => Output của layer là tensor 3 chiều có kích thước: () \*  () \* K
     + Output của convolutional layer sẽ đi qua các hàm kích hoạt trước khi trở thành input của convolutional layer kia
  2. Lớp kích hoạt phi tuyến tính (Rectified Linear Unit)
     + Lớp này được xây dựng với ý nghĩa đảm bảo tính phi tuyến của mô hình huấn luyện sau khi đã thực hiện một loạt các phép tính toán tuyến tính qua các lớp Tích chập. Lớp Kích hoạt phi tuyến nói chung sử dụng các hàm kích hoạt phi tuyến như ReLU hoặc sigmoid, tanh… để giới hạn phạm vi biên độ cho phép của giá trị đầu ra.
     + Phi tuyến tính có nghĩa là đầu ra không thể được sao chép từ tổ hợp tuyến tính của các đầu vào, nếu không có các hàm kích hoạt phi tuyến, thì mạng nơ-ron của chúng ta dù có nhiều lớp vẫn sẽ có hiệu quả như một lớp tuyến tính mà thôi
  3. Lớp tổng hợp (Pooling):
     + có chức năng tổng hợp các bộ lọc lại với nhau, pooling sẽ gom 1 tập các điểm ảnh và cho ra kích thước nhỏ hơn, làm giảm dần kích thước không gian để giảm số lượng tham số và tính toán của ảnh. Ví dụ maxpooling 2x2 thì sẽ chọn ra từ ảnh đầu vào giá trị điểm ảnh lớn nhất trong các nhóm điểm ảnh với kích thước 2x2.
     + Lớp lấy mẫu cũng sử dụng một cửa sổ trượt để quét toàn bộ các vùng trong ảnh tương tự như lớp Tích chập, và thực hiện phép lấy mẫu thay vì phép tích chập – tức là ta sẽ chọn lưu lại một giá trị duy nhất đại diện cho toàn bộ thông tin của vùng ảnh đó.
  4. Lớp kết nối đầy đủ (Fully-connected):
     + tất cả các điểm ảnh được kết nối đầy đủ với node trong lớp tiếp theo
     + output của layer cuối cùng sẽ được là phẳng thành vector và đưa vào một lớp được kết nối như một mạng nơ-ron.
     + Với FC layer được kết hợp với các tính năng lại với nhau để tạo ra một mô hình. Cuối cùng sử dụng softmax hoặc sigmoid để phân loại đầu ra.