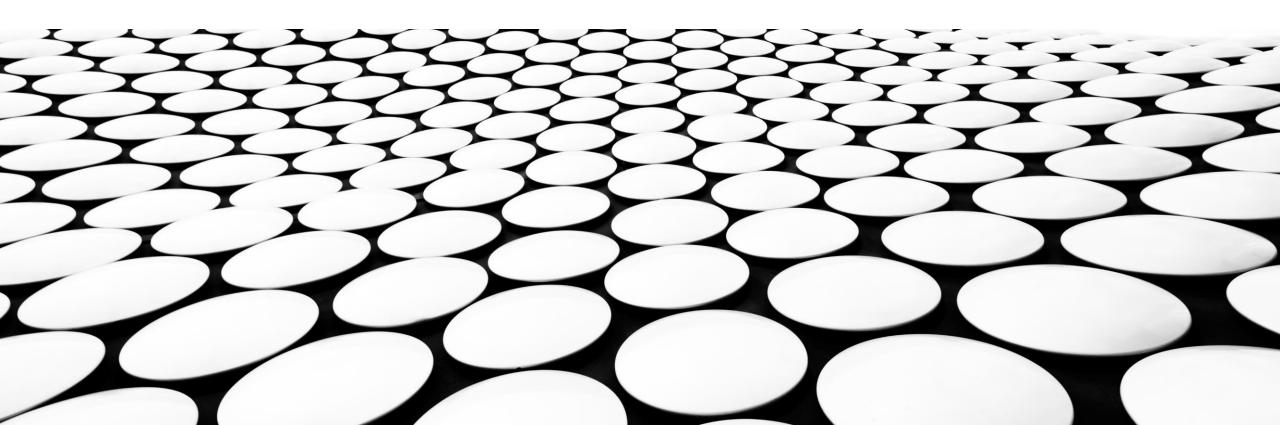
ALGORITHIS

SUPERVISED LEARNING & UNSUPERVISED LEARNING



CÁC THUẬT TOÁN NHÓM NGHIÊN CỨU:

. K nearest neighbor

. K-means for clustering

|. Singular Value | Decomposition

. Convolutional neural network

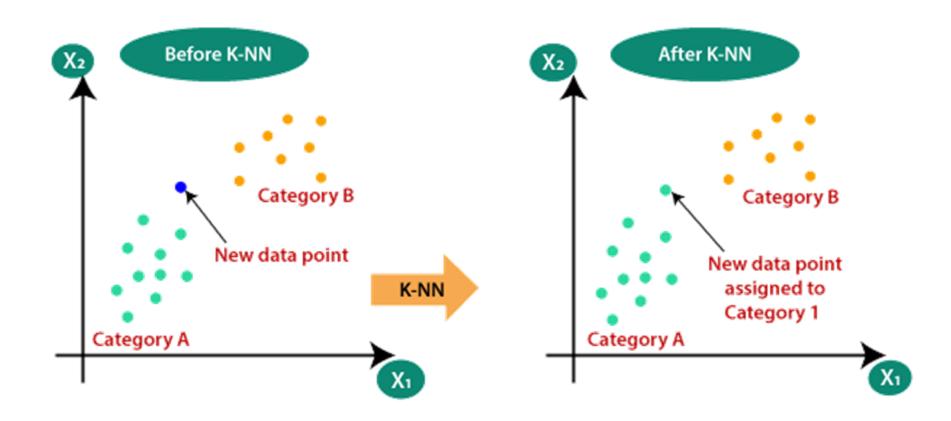
PHÂN BIỆT HỌC CÓ GIÁM SÁT & HỌC KHÔNG GIÁM SÁT

Supervised Learning	Unsupervised Learning		
Dữ liệu để huấn luyện mô hình			
Dữ liệu đầu vào được gắn nhãn.	Dữ liệu đầu vào không được gắn nhãn.		
Chức năng chính			
Classification, Regression	Clustering		
Độ chính xác của kết quả			
Chính xác và đáng tin cậy hơn.	Không chính xác và không đáng tin cậy.		

1. Giới thiệu bài toán:

- KNN là một kĩ thuật máy học phân loại được sử dụng rộng rãi.
- KNN dựa vào số lượng những người hàng xóm gần nhất và nó được sử dụng cho các bài toán hồi quy.

2. Cách thức hoạt động của KNN:



2. Cách thức hoạt động của KNN:

- + Bước 1: Chọn số lượng K neighbors.
- + Bước 2: Lấy K neighbors của điểm data mới theo phương pháp *tính khoảng cách của hai điểm*.
- + Bước 3: Trong K neighbors đó, đếm số lượng data của mỗi cụm.
- + Bước 4: Cập nhật cho cái điểm data mới vào cụm mà có số lượng neighbors vừa đếm ở bước 3 là nhiều nhất.

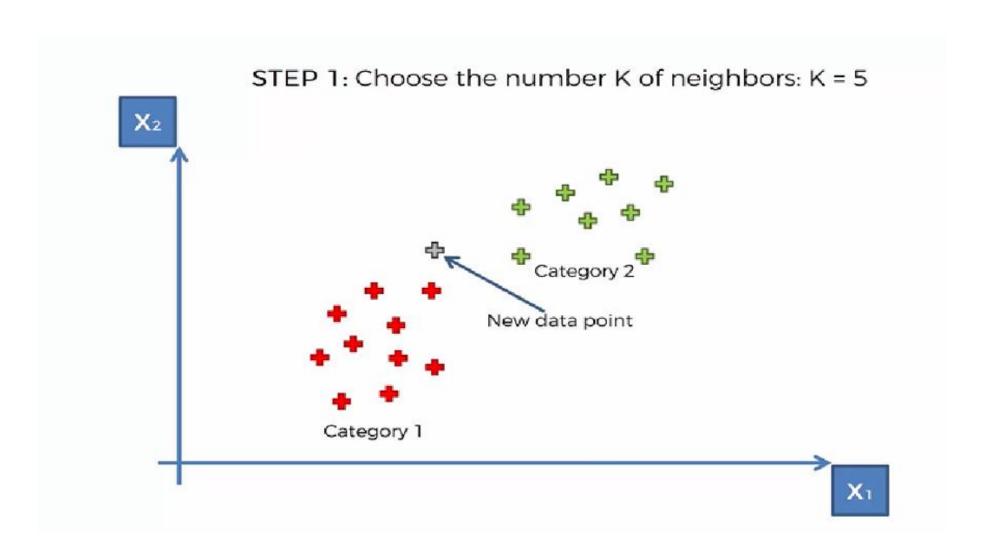
- 3. Công thức tính khoảng cách:
- a) Euclidean distance:
- Là độ dài khoảng cách nối 2 điểm trong không gian
- Công thức:

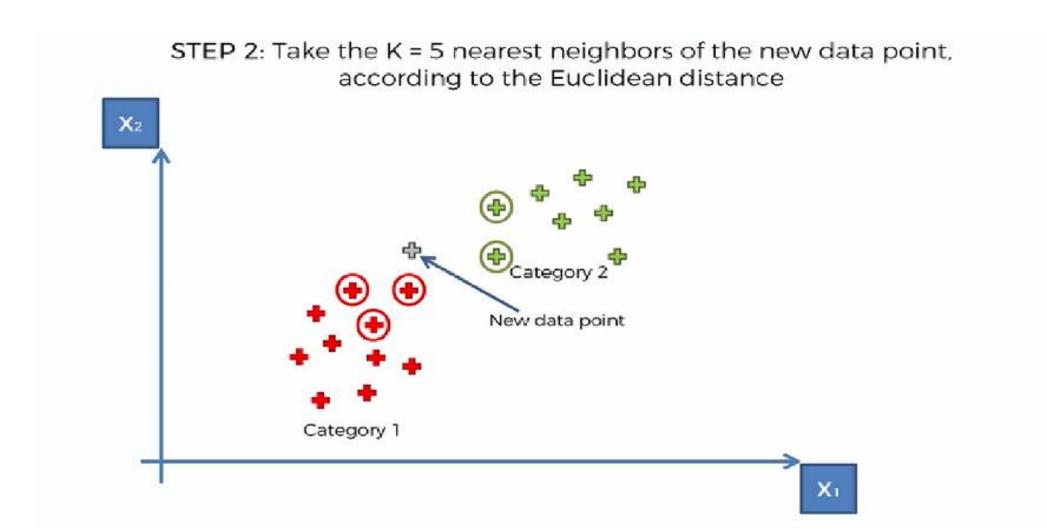
$$d(x, y) = \sqrt{\sum_{i=1}^{n} (y_i - x_i)^2}$$

3. Công thức tính khoảng cách:

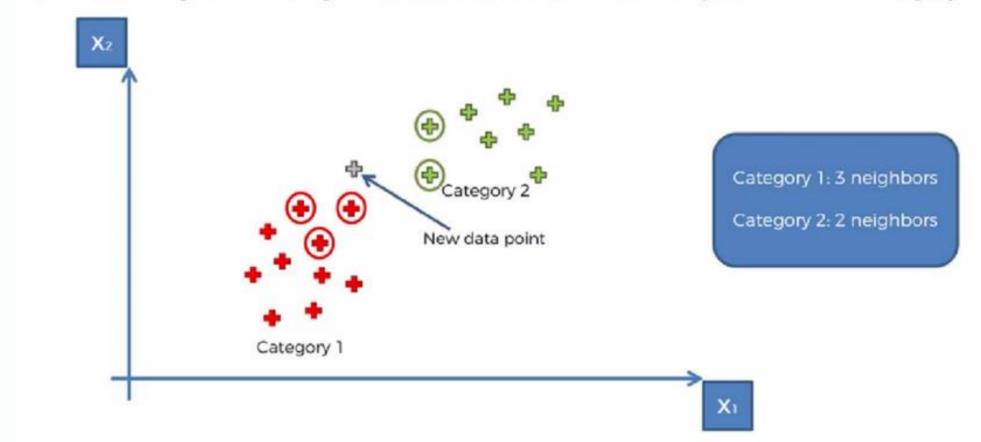
- b) Manhattan distance
- Là khoảng cách L1, một dạng khoảng cách giữa hai điểm trong không gian Euclid với hệ tọa độ Descartes
- Công thức:

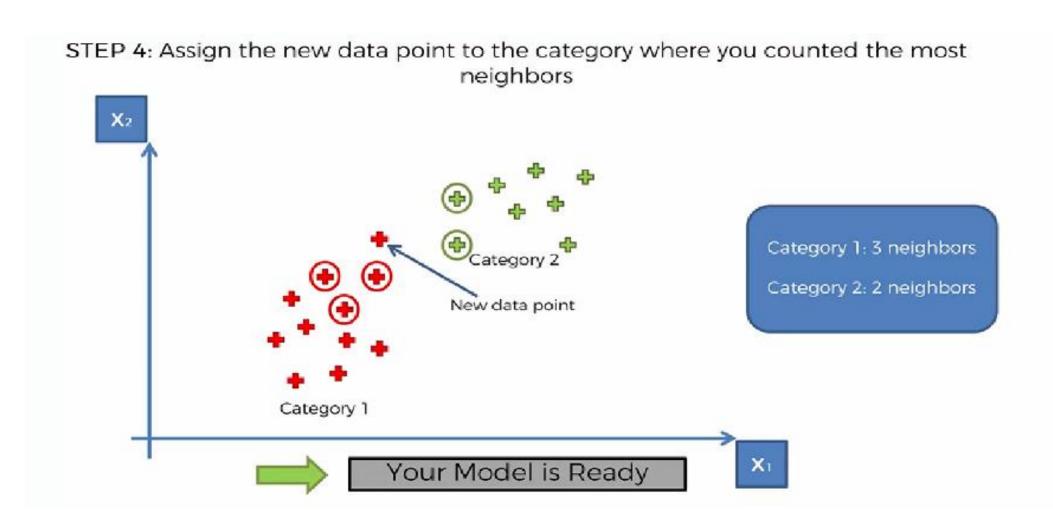
$$d(x, y) = \sum_{i=1}^{n} |x_i - y_i|$$





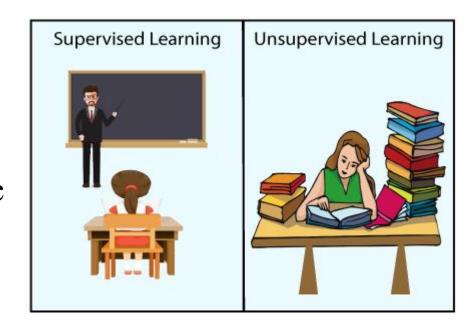
STEP 3: Among these K neighbors, count the number of data points in each category





1. Giới thiệu bài toán:

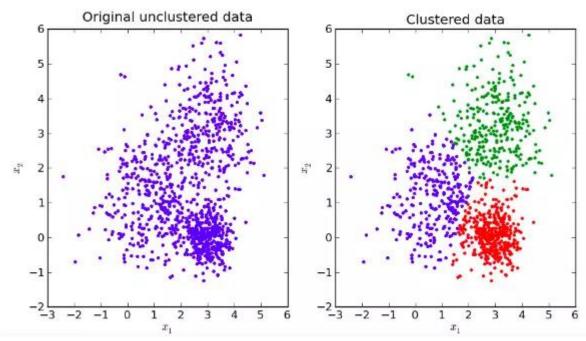
- K-means là thuật toán học không giám sát (Unsupervised Learning).
- Mục tiêu: Phân chia dữ liệu (Không được gán nhãn) thành các cụm sao cho dữ liệu trong cùng 1 cụm có tính chất giống nhau.



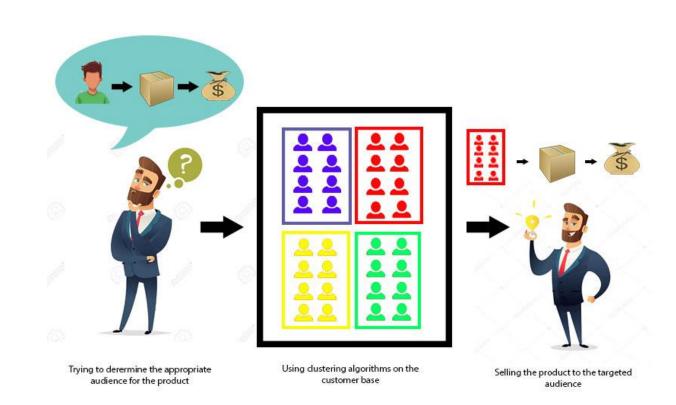
- Mô hình hóa toán học : Dữ liệu đầu vào là X(Không có gắn nhãn) . Đầu ra là K cụm và nhãn cho từng điểm dữ liệu.

1. Giới thiệu về thuật toán:

- Ý tưởng về cụm: Tập hợp các điểm ở gần nhau trong cùng một không gian.
- Ý tưởng về K-means: tìm các điểm centers(điểm đại diện) -> tính khoảng cách của các điểm lân cận tới điểm centers -> so sánh khoảng cách -> cập nhật cụm cho các điểm lân cận.



- Úng dụng thực tế:
- Phân nhóm khách hàng trong kinh doanh.
- Gán nhãn các thuật toán học có giám sát.



2. Tóm tắt cơ sở toán học:

- Input: N điểm dữ liệu X = [x1,x2,...,xN], k là số cluster mong muốn.
- Output: center của mỗi cluster(m1,m2,...,mk) và label vector của mỗi điểm dữ liệu.
- Phân tích thuật toán:
 - + Độ phức tạp: Thấp -> Tương đối
 - + Ưu điểm: Độ phực tạp vừa phải, độ chính xác tương đối cao.
 - + Nhược điểm: Không phù hợp với non-covex data, dễ bị Outliers gây nhiễu, phụ thuộc vào số lượng cluster cần chia.

2. Tóm tắt cơ sở toán học

- Với mỗi điểm dữ liệu x_i , đặt y_ij là vector one-hot của nó: $y_{ij} \in \{0,1\} \ \forall j; \ \sum_{j=1}^n y_{ij} = 1$
- Hàm sai số giữa điểm dữ liệu và center của cluster: $\|\mathbf{x}_i \mathbf{m}_k\|_2^2$
- Khi x_i được phân cụm vào k thì hàm sai số giữa điểm và center được viết lại như sau:

$$\|y_{ik}\|\mathbf{x}_i - \mathbf{m}_k\|_2^2 = \sum_{j=1}^K y_{ij}\|\mathbf{x}_i - \mathbf{m}_j\|_2^2$$

2. Tóm tắt cơ sở toán học:

- Hàm mất mát:

$$\mathcal{L}(\mathbf{Y}, \mathbf{M}) = \sum_{i=1}^N \sum_{j=1}^N y_{ij} \|\mathbf{x}_i - \mathbf{m}_j\|_2^2$$

- Tối ưu hàm mất : Cố định M tìm Y
- Tìm label cho từng điểm dữ liệu: $\mathbf{y}_i = \arg\min_{\mathbf{y}_i} \sum_{j=1}^{n} y_{ij} \|\mathbf{x}_i \mathbf{m}_j\|_2^2$ (3)
- Có thể thu gọn lại tìm label cho từng điểm dữ liệu như sau : $j = rg \min_i \|\mathbf{x}_i \mathbf{m}_j\|_2^2$

2. Tóm tắt cơ sở toán học:

- Tối ưu hàm mất : Cố định Y tìm M
- Cập nhật center cho từng cụm K: $\mathbf{m}_j = \arg\min_{\mathbf{m}_j} \sum_{i=1}^N y_{ij} \|\mathbf{x}_i \mathbf{m}_j\|_2^2$.

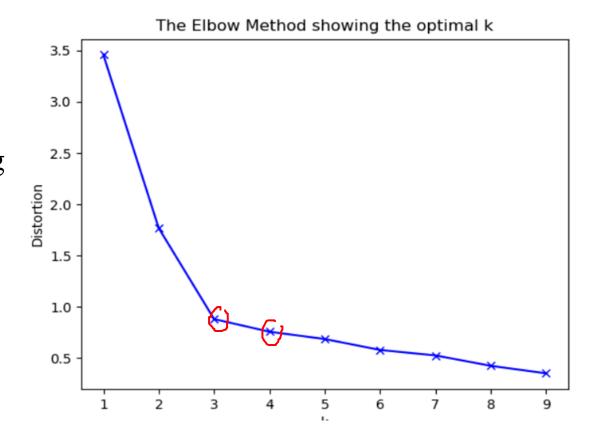
- Đạo hàm để tìm min thì ta sẽ công thức cập nhật center như sau : $\mathbf{m}_j = \frac{\sum_{i=1}^N y_{ij}\mathbf{x}_i}{\sum_{i=1}^N y_{ij}}$

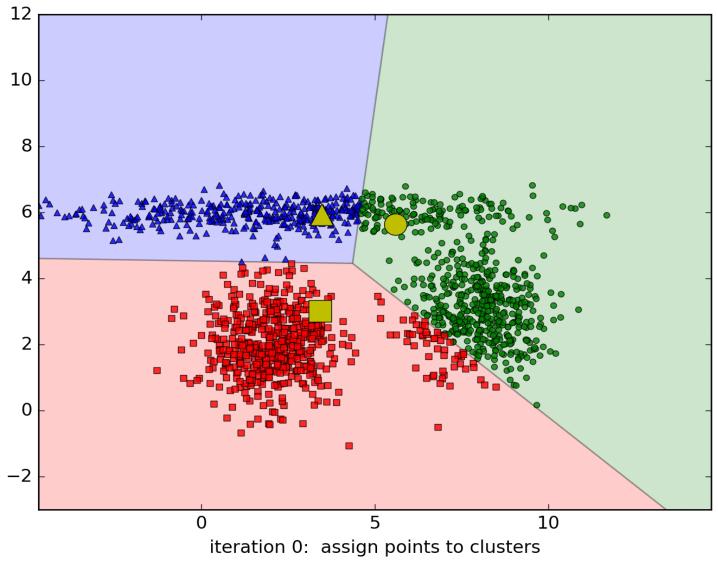
3. Tóm tắt thuật toán:

- INPUT: Dữ liệu X và số lượng cluster cần tìm K.
- OUTPUT: Các center M và label vector cho từng điểm dữ liệu Y.
- 1. Chọn K điểm bất kỳ làm các center ban đầu.
- 2. Phân mỗi điểm dữ liệu vào cluster có center gần nó nhất.
- 3. Nếu việc gán dữ liệu vào từng cluster ở bước 2 không thay đổi so với vòng lặp trước nó thì ta dừng thuật toán.
- 4. Cập nhật center cho từng cluster bằng cách lấy trung bình cộng của tất các các điểm dữ liệu đã được gán vào cluster đó sau bước 2.
- 5. Quay lại bước 2.

4. Đánh giá bài toán K-means bằng Elbow graph:

- Mục đích: Chọn số cluster K phù hợp.
- Cấu trúc Graph:
- Trục hoành : Số lượng cluster.
- Trục trung : Trung bình cộng bình phương khoảng cách giữa tâm cluster đén các điểm còn lại.





1. Giới thiệu thuật toán:

- Phương pháp SVD đã được phát triển dựa trên những tính chất của ma trận trực giao và ma trận đường chéo để tìm ra một ma trận xấp xỉ với ma trận gốc.

- 2. Một số kiến thức về đại số tuyến tính:
- a) Trị riêng và vector riêng
- Cho một ma trận vuông A cấp n.
- Số λ được gọi là *trị riêng* của A nếu tồn tại véctơ $x \in \mathbb{R}^n$, $x \neq \theta$ thoả mãn:

$$Ax = \lambda x$$

Khi đó vector $x \neq \theta$ được gọi là *vécto riêng* của A ứng với trị riêng λ .

- 2. Một số kiến thức về đại số tuyến tính:
- b) Hệ vectơ trực giao, trực chuẩn
- Xét hệ cơ sở $\mathbf{u} = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, ..., \mathbf{u}_{\mathbf{m}}\}$ trong $\mathbb{R}^{\mathbf{m}}$.
- Nếu u thỏa: $\begin{cases} u_i \neq \theta \\ \left\langle u_i, u_j \right\rangle = 0 \ \forall \ 1 \leq i,j \leq m \ ; \ i \neq j \end{cases} ^{(*)}$

Thì khi đó u là một hệ trực giao.

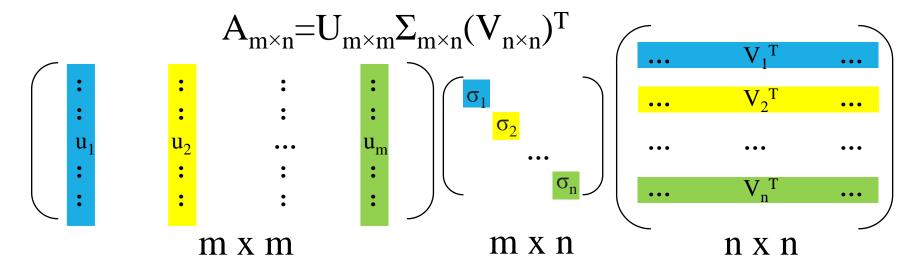
• Nếu u thỏa (*) và chuẩn (norm) $\|\mathbf{u}_{\mathbf{i}}\|_2 = 1$. Khi đó u là một hệ trực chuẩn.

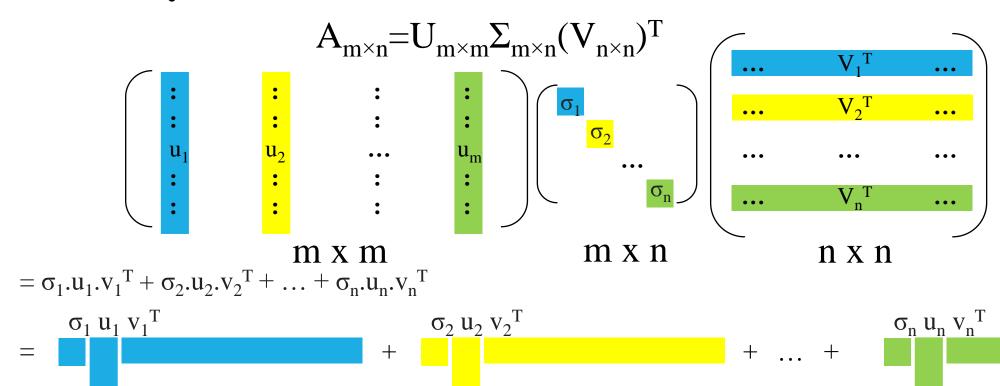
1	2	3	4
2	4	6	8
10	20	30	40
20	40	60	80

	1
	2
=	10
	20

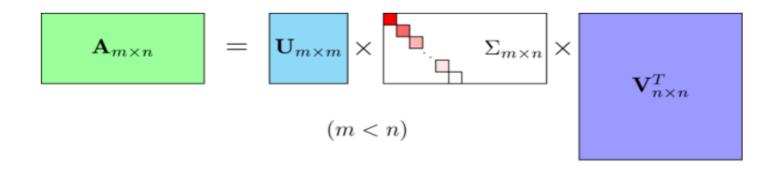
\otimes	1	2	3	4

1	2	3	4
5	6	99	55
23	255	120	34
75	42	60	0





$$A_{m \times n} = U_{m \times m} \Sigma_{m \times n} (V_{n \times n})^T$$



$$= \begin{array}{c} \\ \mathbf{A}_{m \times n} \end{array} \times \begin{array}{c} \\ \mathbf{V}_{n \times n} \\ \\ \Sigma_{m \times n} \end{array} \times \begin{array}{c} \mathbf{V}_{n \times n}^T \\ \\ \end{array}$$

3. Thuật toán SVD:

Tính U, Σ , V

- Bước 1: Tính A^T A
- Bước 2: Tìm trị riêng và vectơ riêng của A^TA
 - + $(A^{T}A)x = \lambda x \Leftrightarrow (A^{T}A \lambda I)x = 0$
 - + T = $(A^TA \lambda I)$. Giải $det(T) = 0 \rightarrow \lambda = ????$
- Bước 3: Tìm hệ vectơ trực giao từ λ thay vào T → Tìm được hệ trực chuẩn → V

3. Thuật toán SVD:

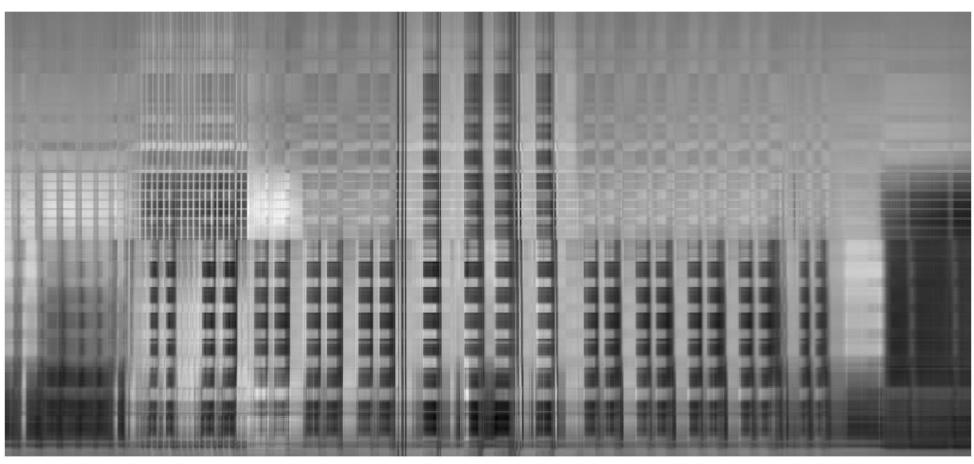
Tính U, Σ , V

- Bước 4: Tìm Σ
 - + Điền các giá trị căn bậc hai các λ tìm được vào ma trận theo đường chéo, giảm dần trên ma trận.
- Bước 5: Tìm U

$$+ A = U\Sigma V^{T} \Leftrightarrow AV = US \Leftrightarrow AVS^{-1} = U \rightarrow U$$

$$U,\Sigma,V \Longrightarrow A' = U'_{mxk} \Sigma'_{kxk} (V'_{kxn})^T (\cong A)$$

k = 5: error = 0.1492



TỔNG QUAN

- Lịch Sử
- Tổng quan về cấu trúc trong CNN
- Ví dụ về cách thức hoạt động của CNN và giới thiệu về các thuật ngữ quan trọng
- CNNs for NLP

LỊCH SỬ

Yann LeCun

New York University

Facebook Artificial Intelligence Research

Yoshua Bengio

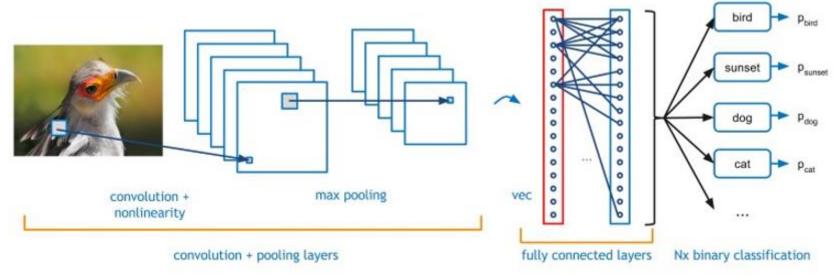
Université de Montréal





Vào năm 1995, Yann LeCun và Yoshua Bengio đã giới thiệu khái niệm về convolutional neural networks.

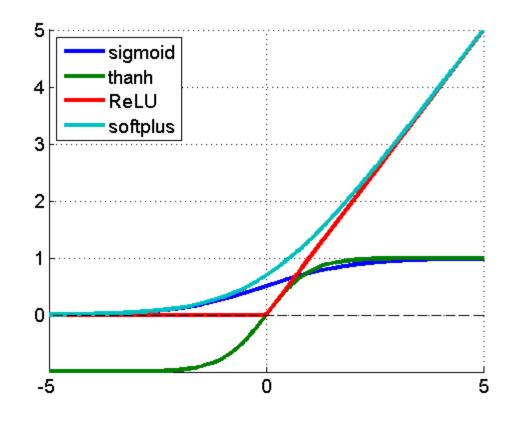
CÁU TRÚC TRONG CNN



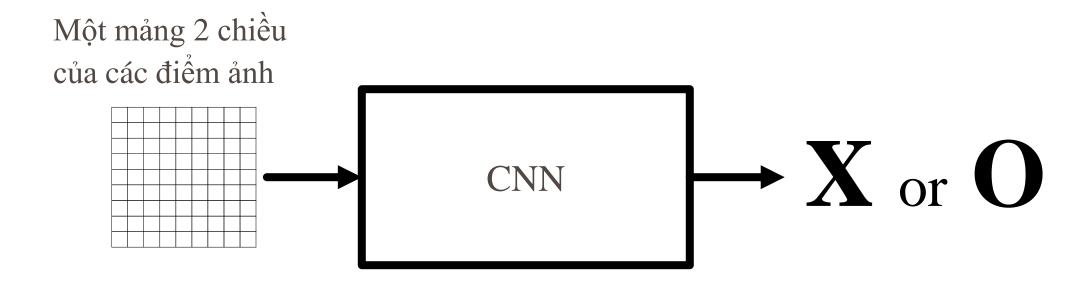
http://www.nallatech.com/fpga-acceleration-convolutional-neural-networks/

NON-LINEARITY?

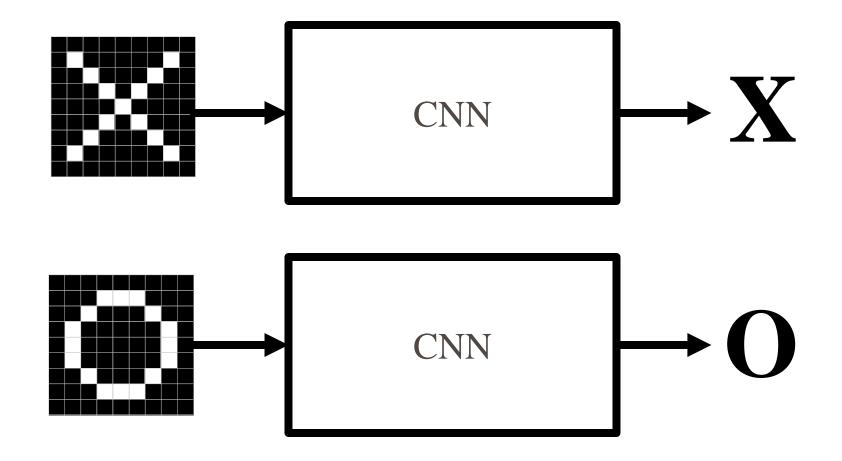
- Một mạng neural chỉ là phi tuyến tính nếu
 'squash' tín hiệu đầu ra từ các node có chức năng
 kích hoạt phi tuyến tính
 - => Xử dụng hàm tích lũy
- Việc diễn giải tín hiệu đầu ra 'squashed' có thể được hiểu là cường độ của tín hiệu này



VÍ DỤ

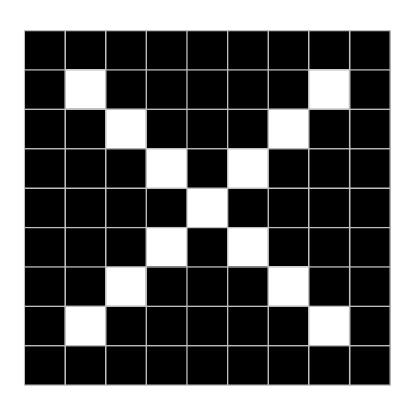


VÍ DỤ

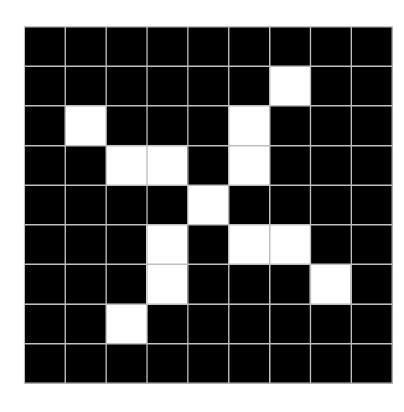


MỘT SỐ HÌNH ẢNH KHÁC rotation weight translation scaling

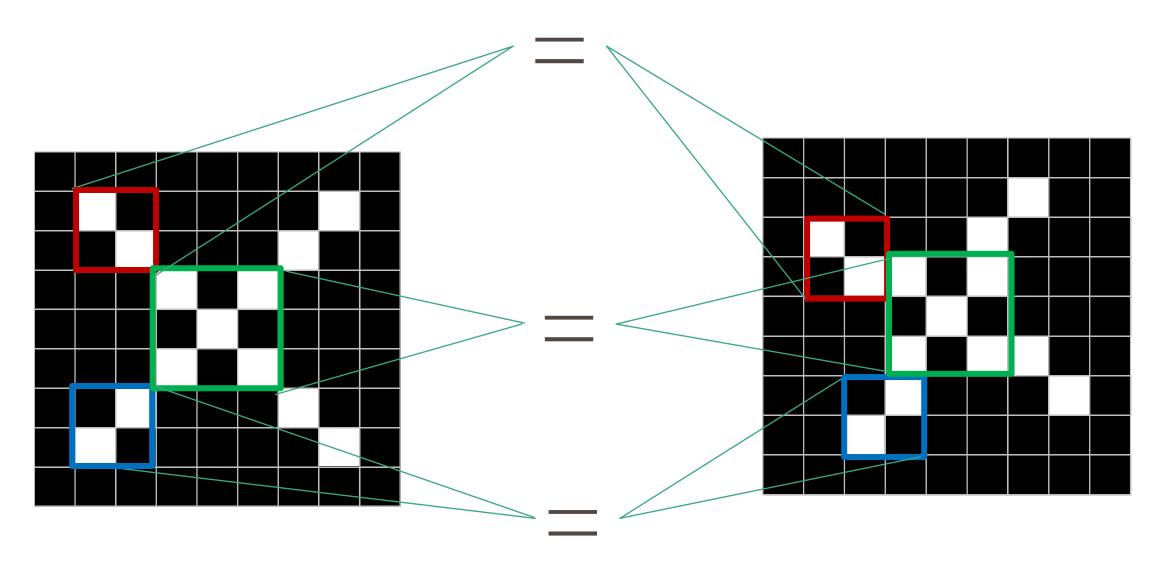
LÀM SAO ĐỂ SO SÁNH?







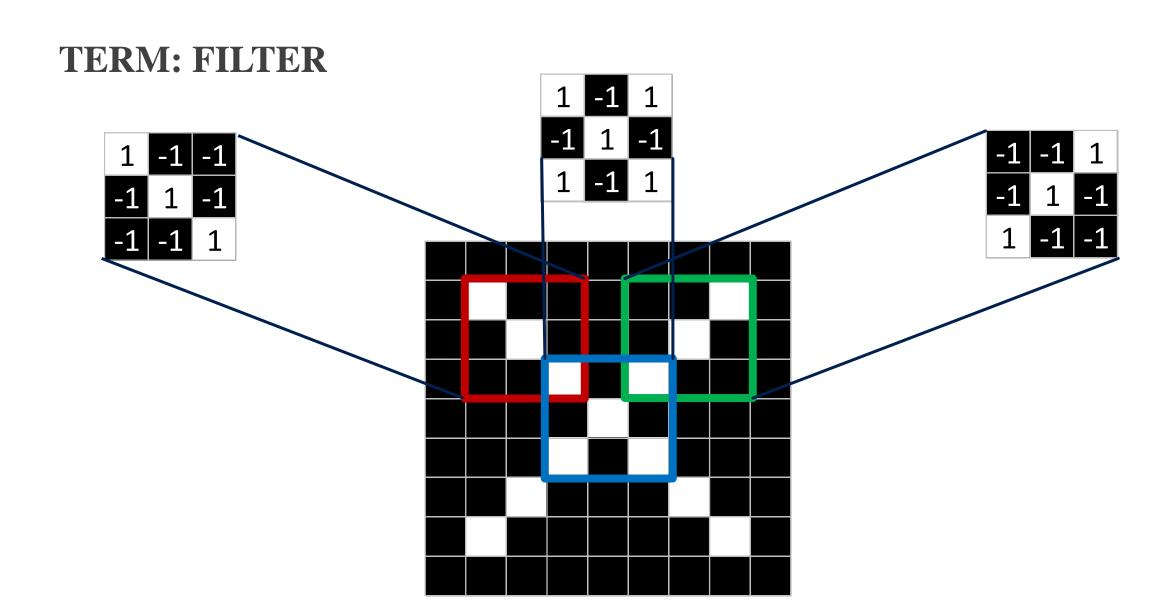
SO SÁNH VÙNG NHỎ



TERM: FILTER

1	-1	-1
-1	1	-1
-1	-1	1

-1	-1	1
-1	1	-1
1	-1	-1



TERM: CONVOLUTION

1,	1 _{×0}	1,	0	0
0,×0	1,	1,0	1	0
0 _{×1}	O _{×0}	1,	1	1
0	0	1	1	0
0	1	1	0	0

4	

Image

Convolved Feature

Các fitlers sử dụng để tạo một filter layer dựa trên dữ liệu đầu vào

- Vùng màu vàng: đếm tất cả những cái nằm trong đường chéo của nó
- Khu vực màu xanh lá cây: các tính năng đầu vào

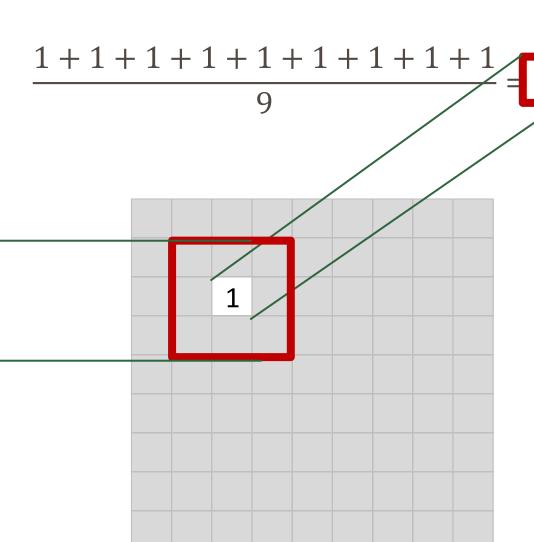
http://deeplearning.stanford.edu/wiki/index.php/Feature_extraction_using_convolution

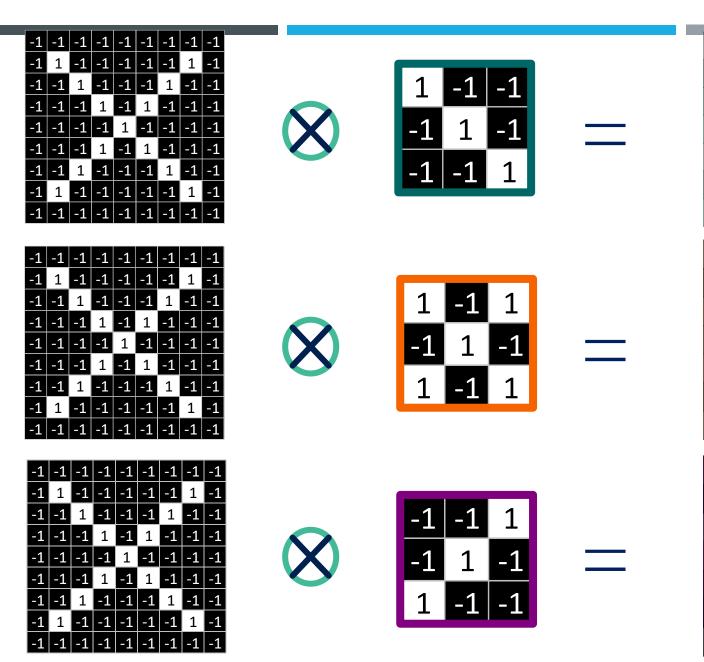
BUILDING THE FILTER LAYER

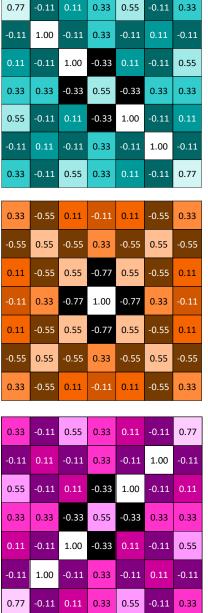
1	-1	-1
-1	1	-1
-1	-1	1
$\overline{}$		

1	1	1
1	1	1
1	1	1

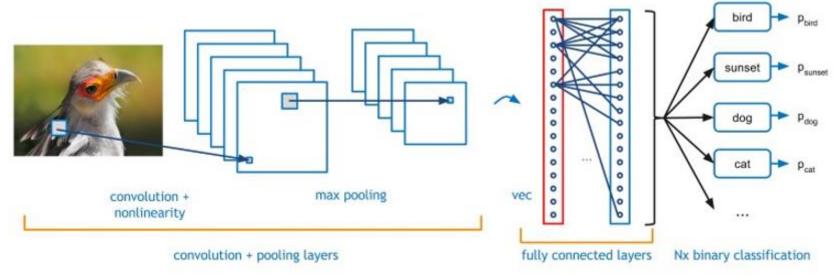
								-1
								-1
-1	-1	1	-1	-1	-1	1	-1	-1
								-1
-1	-1	-1	-1	1	-1	-1	-1	-1
-1	-1	-1	1	-1	1	-1	-1	-1
-1	-1	1	-1	-1	-1	1	-1	-1
-1	1	-1	-1	-1	-1	-1	1	-1
-1	-1	-1	-1	-1	-1	-1	-1	-1







CÁU TRÚC TRONG CNN



http://www.nallatech.com/fpga-acceleration-convolutional-neural-networks/

TERM: POOLING LAYERS

Single depth slice

 1
 1
 2
 4

 5
 6
 7
 8

 3
 2
 1
 0

 1
 2
 3
 4

max pool with 2x2 filters and stride 2

6	8
3	4

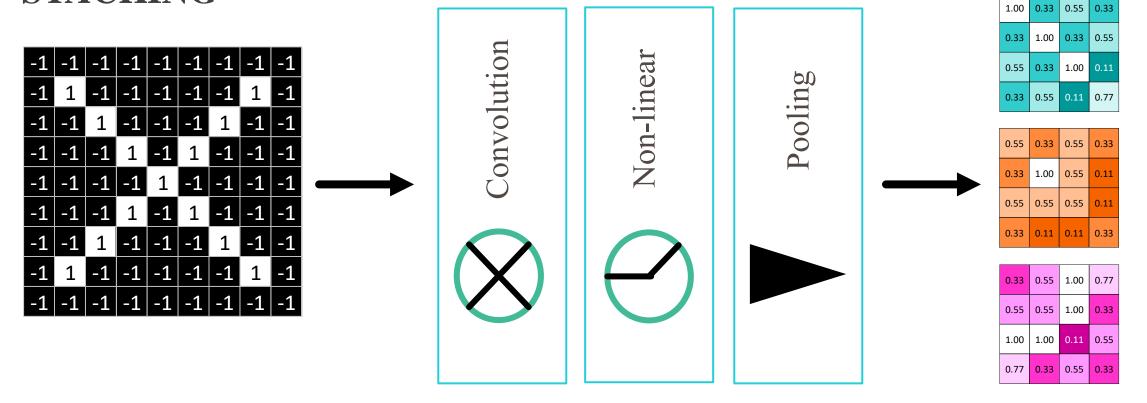
0.77	-0.11	0.11	0.33	0.55	-0.11	0.33
-0.11	1.00	-0.11	0.33	-0.11	0.11	-0.11
0.11	-0.11	1.00	-0.33	0.11	-0.11	0.55
0.33	0.33	-0.33	0.55	-0.33	0.33	0.33
0.55	-0.11	0.11	-0.33	1.00	-0.11	0.11
-0.11	0.11	-0.11	0.33	-0.11	1.00	-0.11
0.33	-0.11	0.55	0.33	0.11	-0.11	0.77
0.33	-0.55	0.11	-0.11	0.11	-0.55	0.33
-0.55	0.55	-0.55	0.33	-0.55	0.55	-0.55
0.11	-0.55	0.55	-0.77	0.55	-0.55	0.11
-0.11	0.33	-0.77	1.00	-0.77	0.33	-0.11
0.11	-0.55	0.55	-0.77	0.55	-0.55	0.11
-0.55	0.55	-0.55	0.33	-0.55	0.55	-0.55
0.33	-0.55	0.11	-0.11	0.11	-0.55	0.33
0.33	-0.11	0.55	0.33	0.11	-0.11	0.77
-0.11	0.11	-0.11	0.33	-0.11	1.00	-0.11
0.55	-0.11	0.11	-0.33	1.00	-0.11	0.11
0.33	0.33	-0.33		-0.33		0.33
0.11	-0.11	1.00	-0.33	0.11	-0.11	0.55
-0.11	1.00	-0.11	0.33	-0.11	0.11	-0.11
0.77	-0.11	0.11	0.33	0.55	-0.11	0.33

1.00	0.33	0.55	0.33
0.33	1.00	0.33	0.55
0.55	0.33	1.00	0.11
0.33	0.55	0.11	0.77

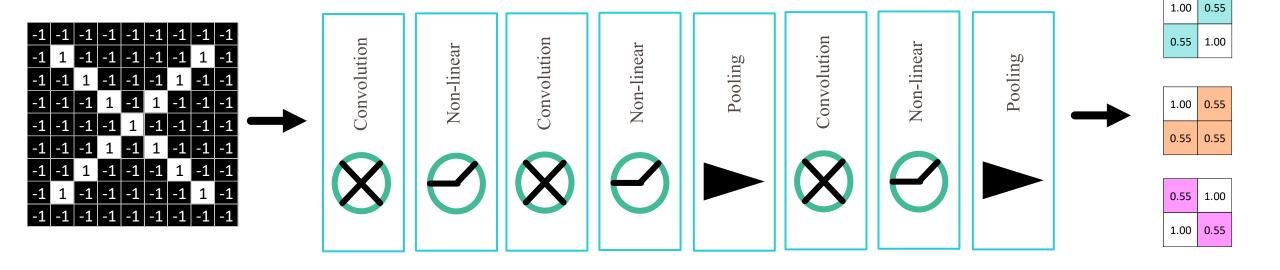
0.55	0.33	0.55	0.33
0.33	1.00	0.55	0.11
0.55	0.55	0.55	0.11
0.33	0.11	0.11	0.33

0.33	0.55	1.00	0.77
0.55	0.55	1.00	0.33
1.00	1.00	0.11	0.55
0.77	0.33	0.55	0.33

STACKING

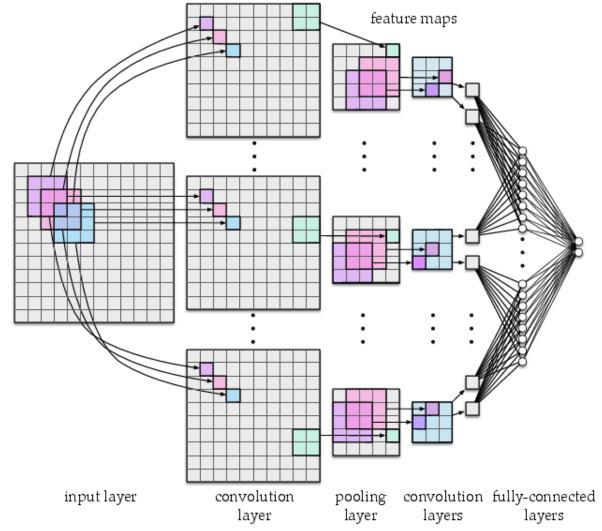


STACKING

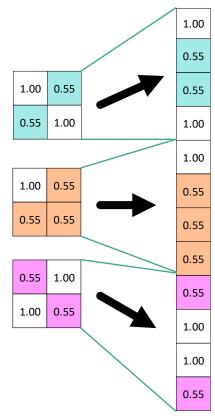


TERM: FULLY-CONNECTED

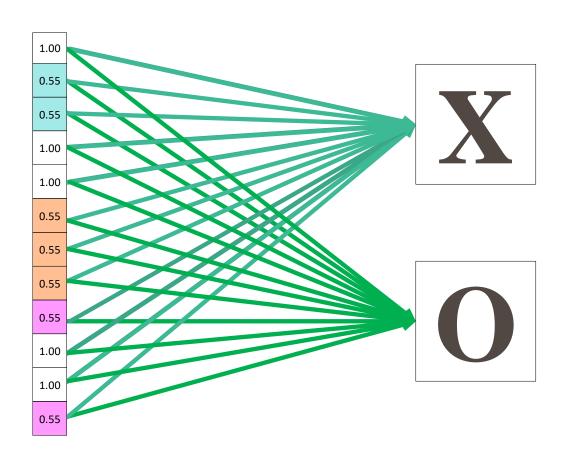
Lấy một số lượng đầu vào và xuất ra một vecto N chiều



VÍ DŲ: FULLY-CONNECTED LAYER



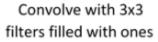
FULLY-CONNECTED LAYER => TRÅ VÈ OUTPUT

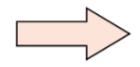


TERM: STRIDES

- Stride là số pixel thay đổi trên ma trận đầu vào. Khi stride là 1 thì ta di chuyển các kernel 1 pixel. Khi stride là 2 thì ta di chuyển các kernel đi 2 pixel và tiếp tục như vậy. Hình dưới là lớp tích chập hoạt động với stride là 2.

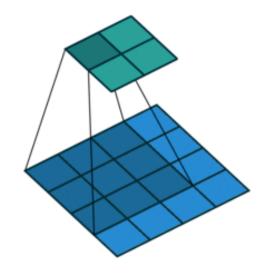
1	2	3	4	5	6	7
11	12	13	14	15	16	17
21	22	23	24	25	26	27
31	32	33	34	35	36	37
41	42	43	44	45	46	47
51	52	53	54	55	56	57
61	62	63	64	65	66	67
71	72	73	74	75	76	77



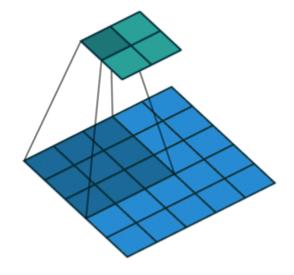


108	126	
288	306	

TERM: STRIDES

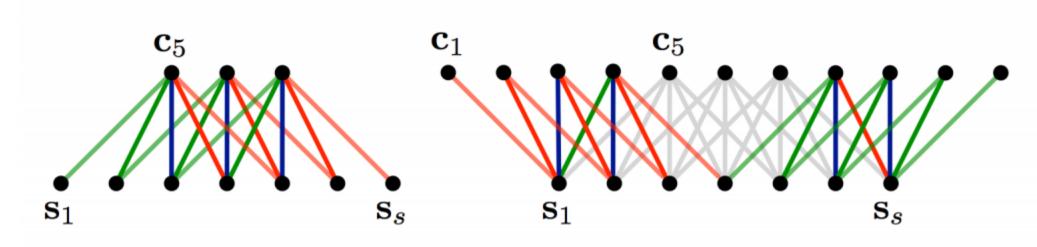


without strides



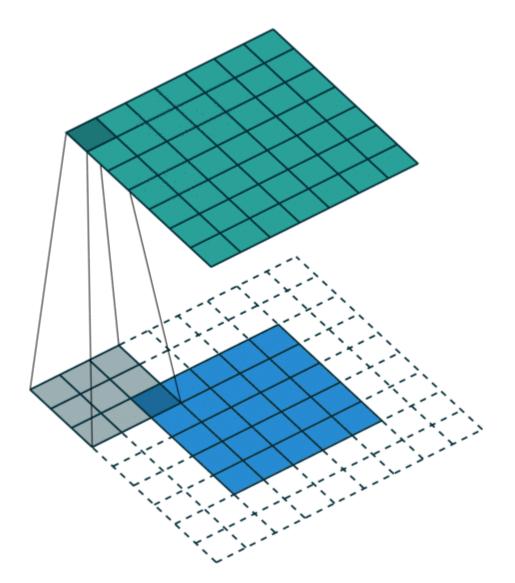
with strides

TERM: PADDING



A Convolutional Neural Network for Modelling Sentences (2014)

TERM: PADDING

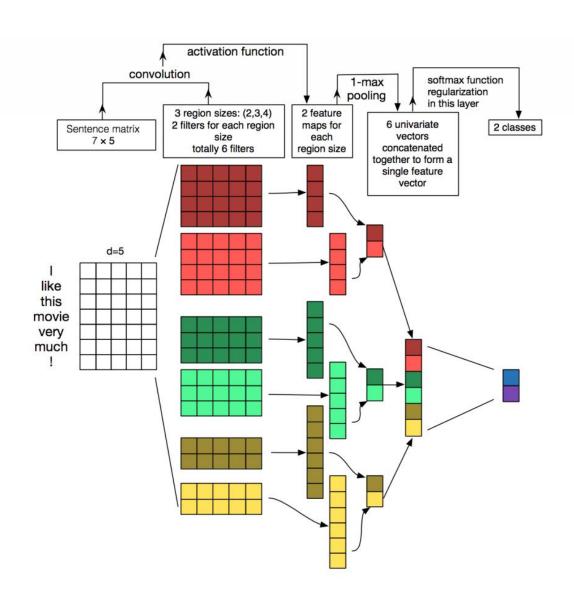


NLP (NEURO LINGUISTIC PROGRAMMING): PHÙ HỢP CHO NHỮNG LĨNH VỰC NÀO?

Phân loại, Ví dụ:

- Phân tích cảm xúc
- Phát hiện các thư rác
- Phân loại các chủ đề

NLP



THANKS FOR LISTENING

Thành viên nhóm:

- 1. Nguyễn Vũ Dương 20520465
- 2. Bùi Hữu Đức 20520449
- 3. Phạm Phước An 20520375
- 4. Hoàng Công Danh 20520431