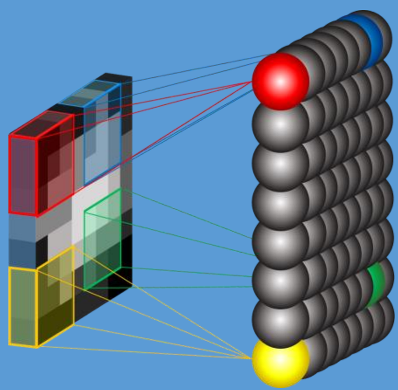
EMS702P Statistical Thinking and Applied Machine Learning

Week 10.1 – CNN and Unsupervised learning

Yunpeng Zhu



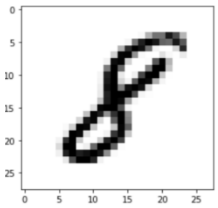
Unsupervised learning

Edition: v1.1

Table of Contents

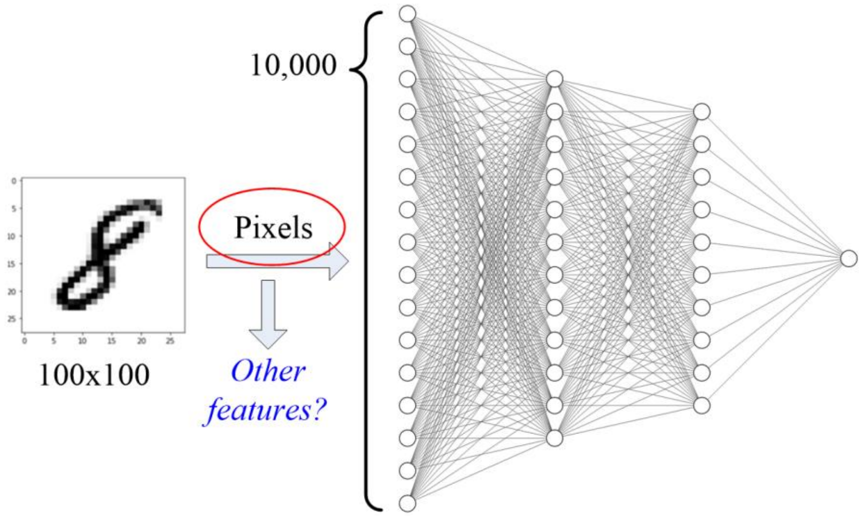
1 Convolutional Neural Networks (CNN)

The CNN was developed to imitate the eye systems of cats, which can be used to solve prediction and classification of images [1].



Consider the specific task of recognizing handwritten digits. Each input image comprises a set of pixel intensity values, and the desired output is a posterior probability distribution over the ten digit classes.

One simple approach would be to treat the image as the input to a fully connected network. Given a sufficiently large training set, such a network could in principle yield a good solution to this problem.

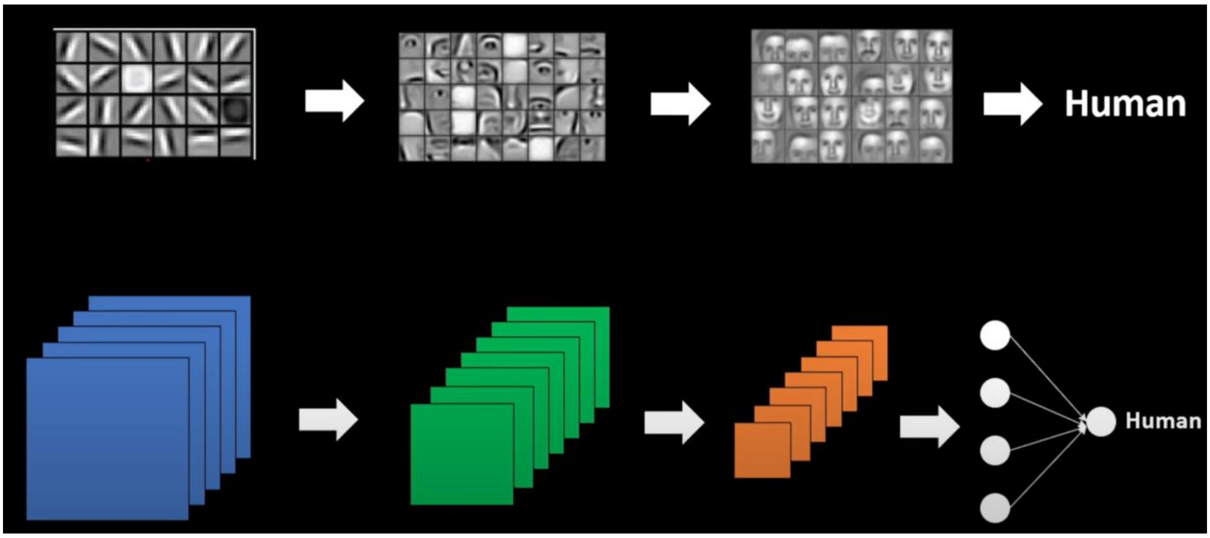


Quiz 1.1: What features can be applied to characterize a picture?

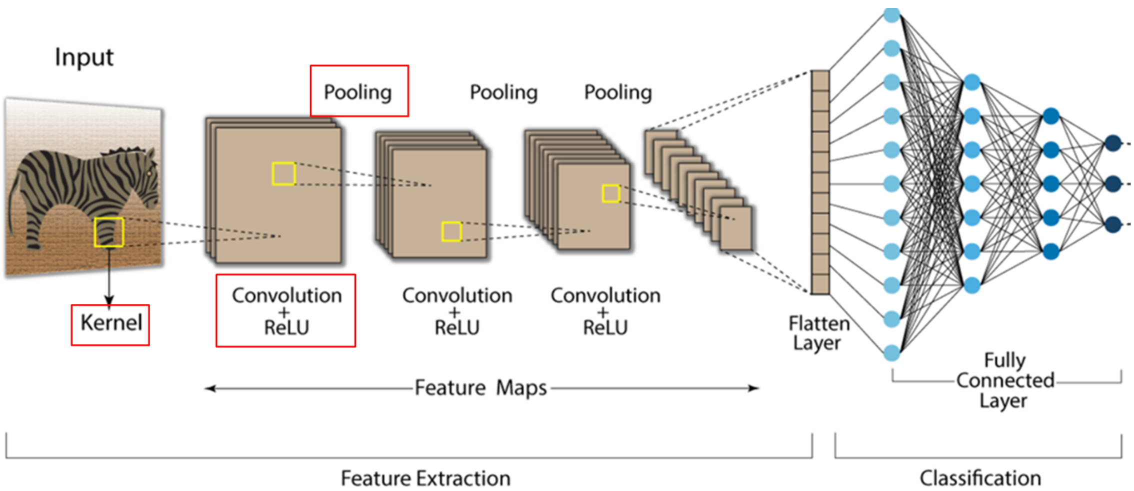
We expect to extract these features from a figure and then use them to identify the plot. **BUT** How could a computer know these features?

Considering we are looking at a very big picture, we can only focus on a small subregions of the image, which is known as the "**receptive field**".

The basic idea of CNN is to imitate this mechanism.



and the basic structure of CNN is



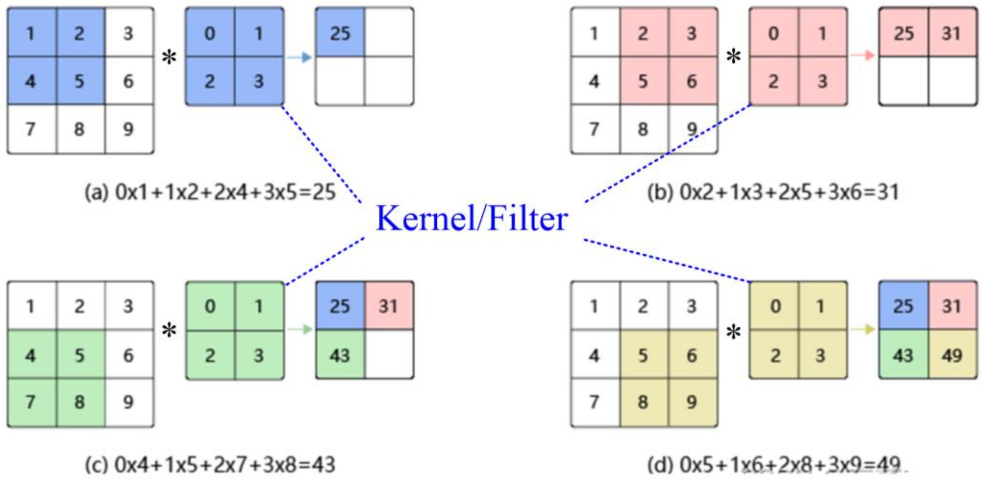
You can find some new terminologies in the above plot:

Kernel, Convolution, Pooling

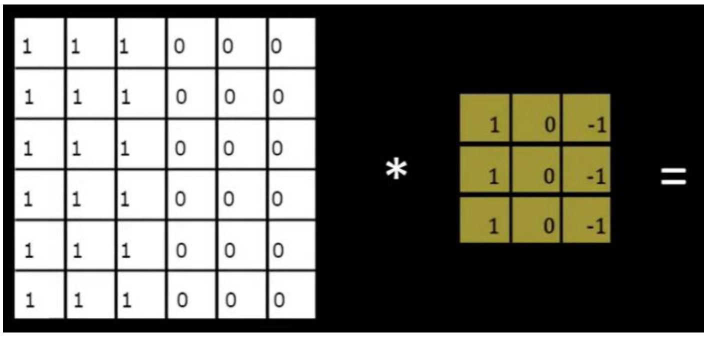
Convolution operation is the heart of CNN.

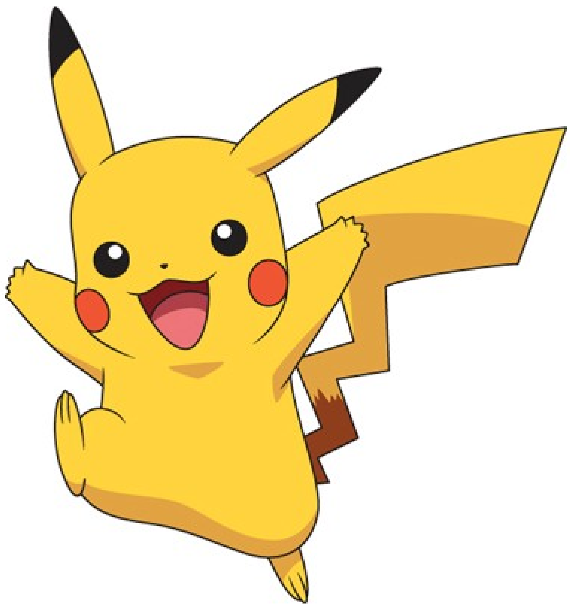
The convolution operation [2]:

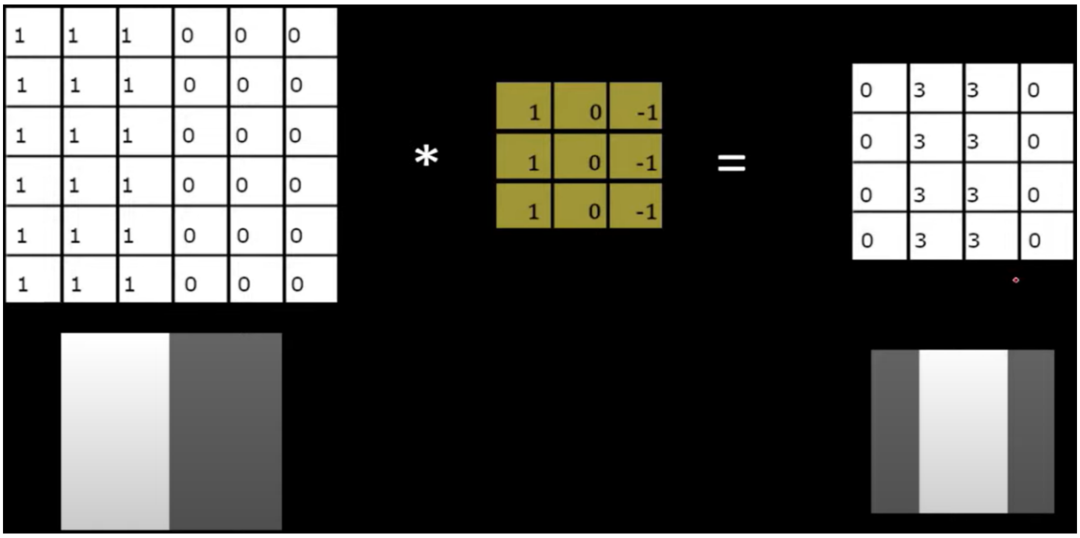
We use “*” to represent the convolution operation.



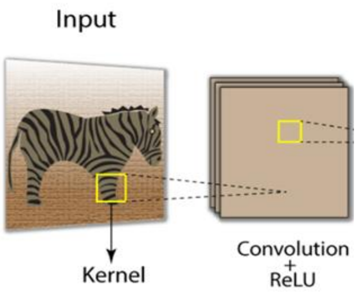
Quiz 1.2: Calculate the convolutions:







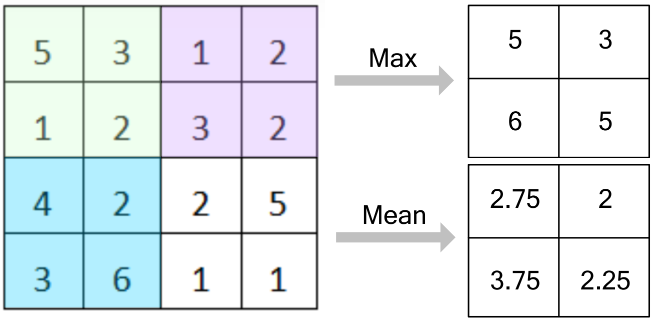
Such filter can find out the **vertical feature** in a picture

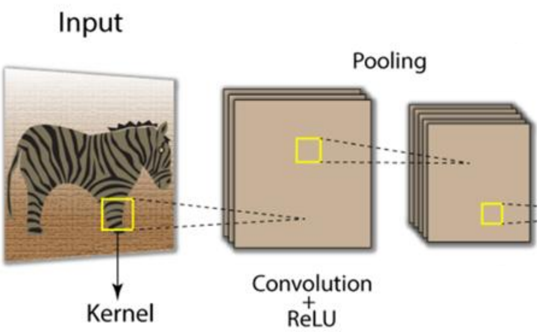


If we have multiple filters that can find out vertical, horizontal, oblique, circle, etc. features, we have multiple layers in the convolution layer of a CNN.

The sub-sampling operation (Pooling):

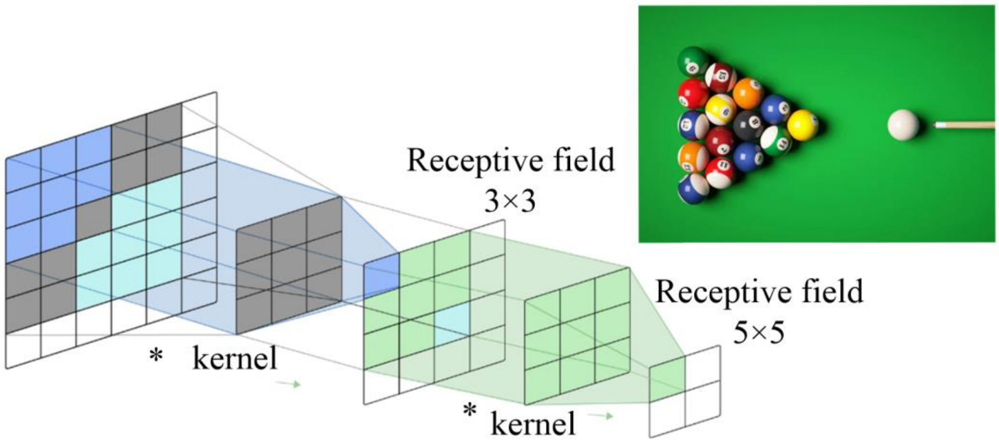
Reduce the dimension of the convolution layer. Typical sub-sampling approaches include maximum and average method.



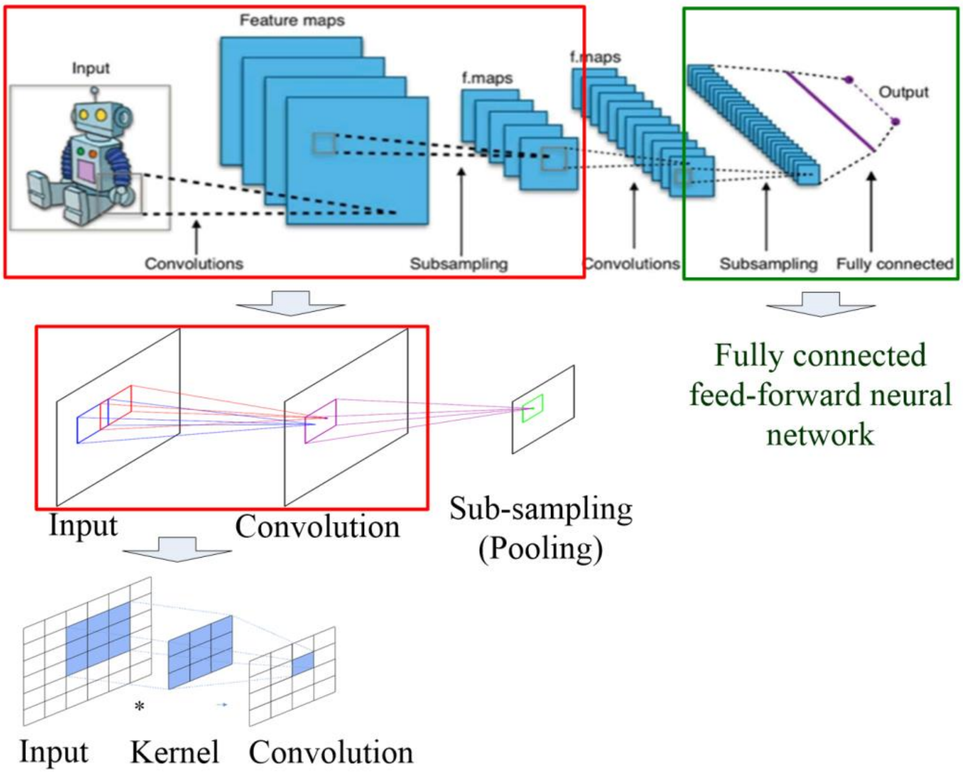


Expanding the receptive field:

Use multiple convolution layers to expand the receptive field



Full CNN:

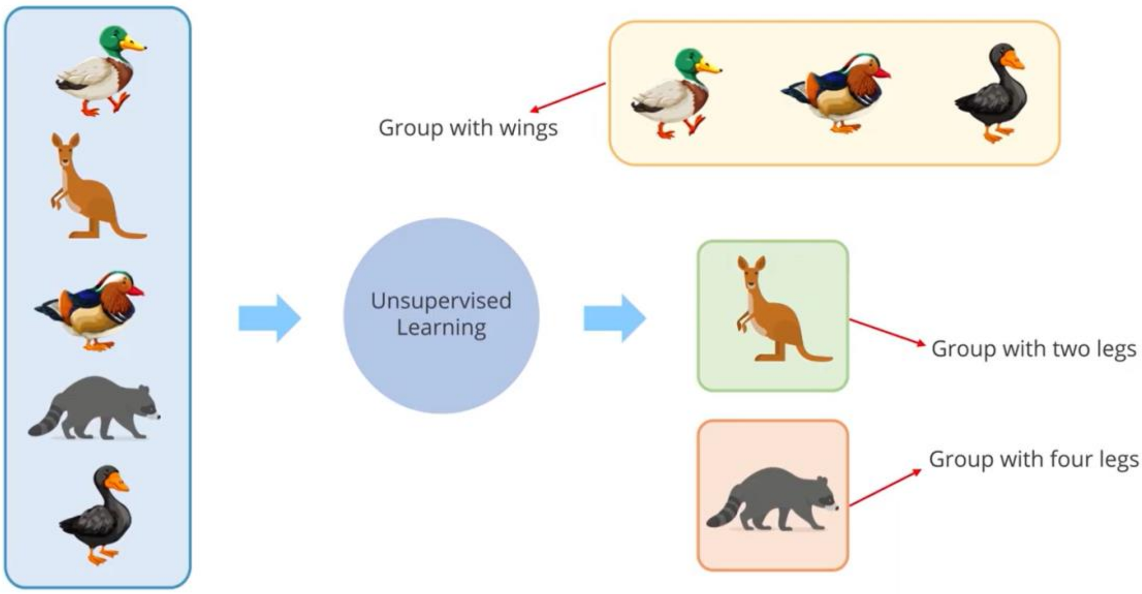


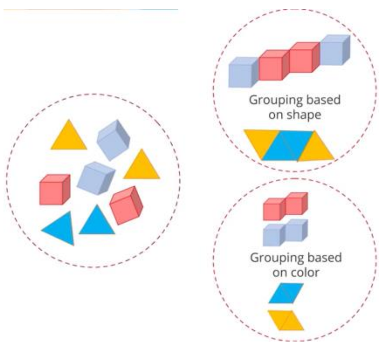
2 Introduction of unsupervised learning

Unsupervised learning is a type of algorithm that learns patterns from unlabeled data. The unlabeled data cannot be directly used for regression and classification as we introduced in the previous lectures.

Supervised learning	Unsupervised learning
Deals with labeled data where the output data patterns are known to the system	works with unlabeled data where the output is just based on the collection of perceptions
● Regression	● Feature extraction
● Classification	● Clustering
Less complex	More complex
Off-line analysis	Real-time analysis

For example, consider you ask children to find birds in the following pictures → They cannot do that because they don't even know what is a bird → But they can always find groups of animals with similar characters.

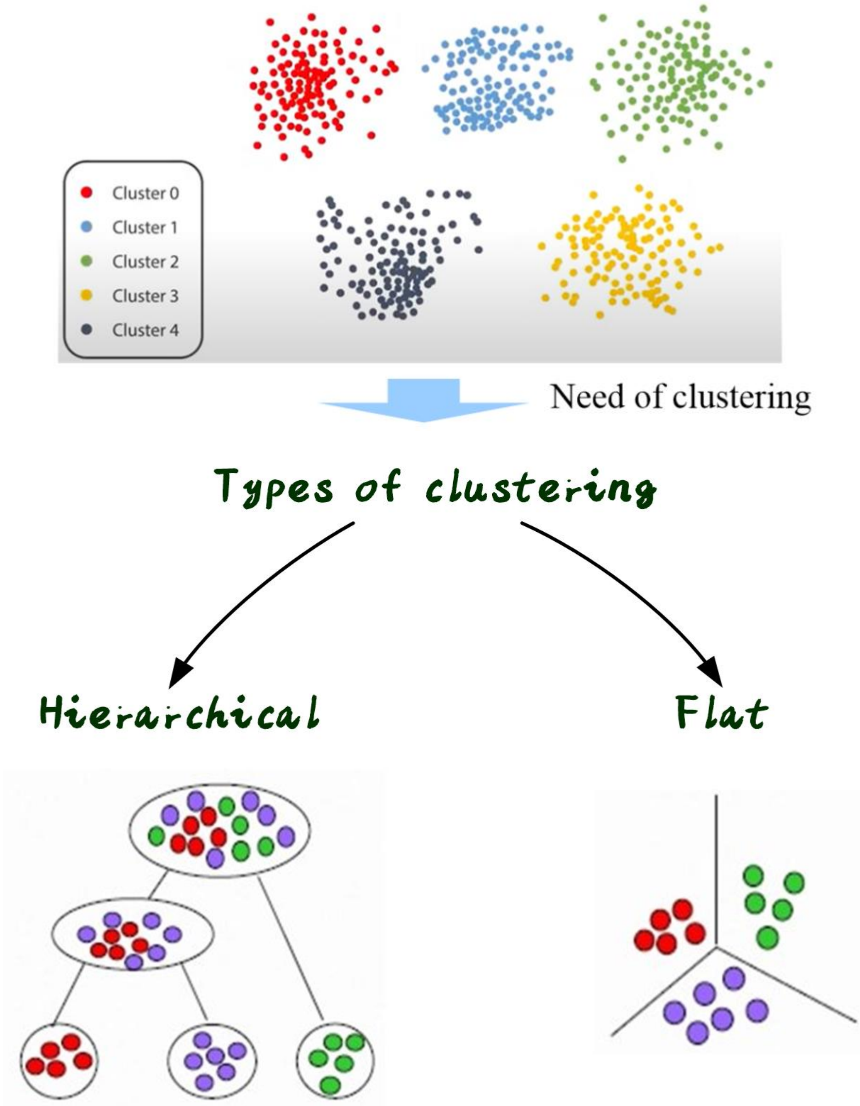




In practice, there is no right or wrong way to cluster data patterns. But which one is better? This will depend on the similarity measurement defined by the users (either color or shape).

3 Clustering

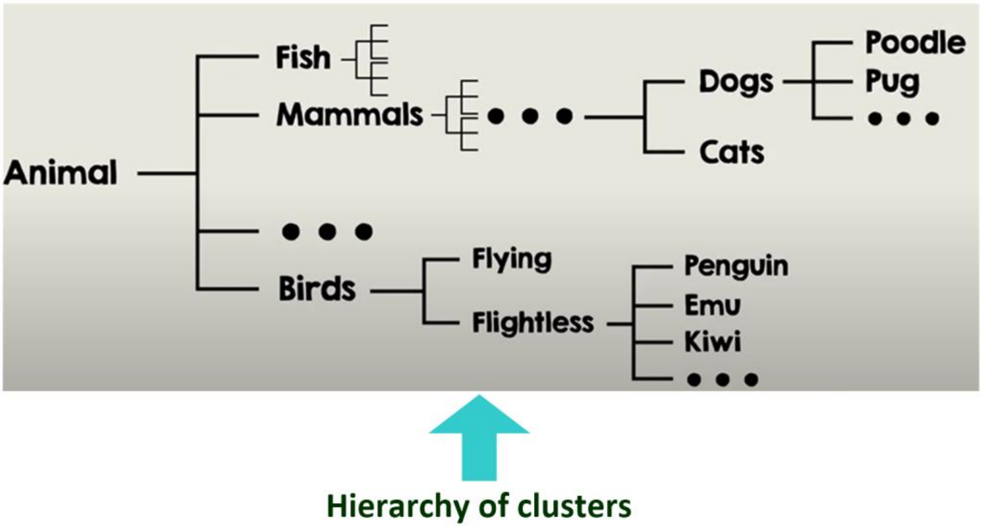
The method of grouping similar entities together is called **clustering**.



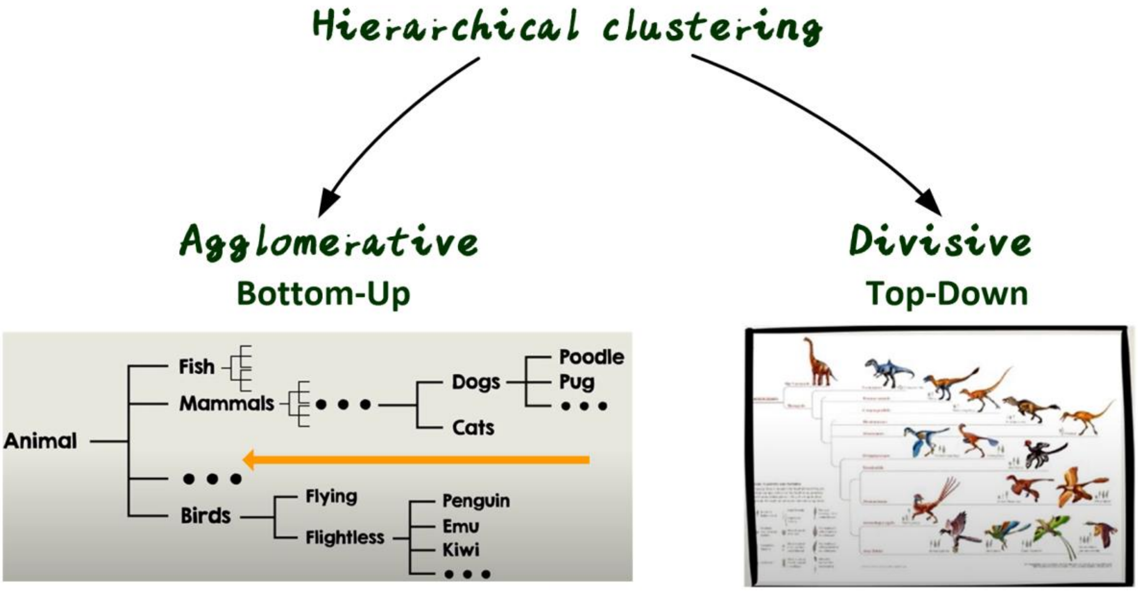
3.1 Hierarchical clustering

Historically, hierarchical clustering **[3]** was developed first, and we can get acquainted with it.

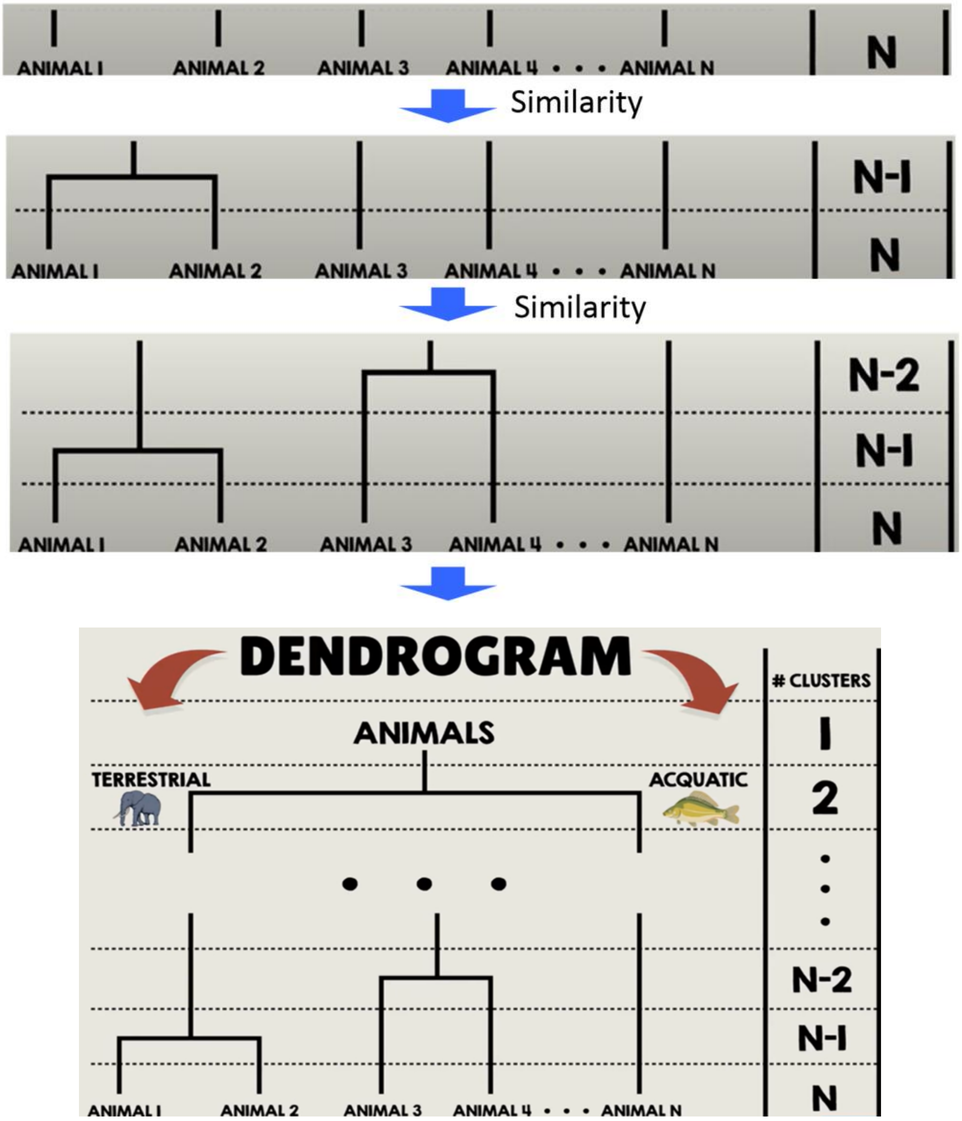
Taxonomy of the animal kingdom



There are **two** types of hierarchical clustering:



Question: In unsupervised learning and clustering, which clustering method is better and why?



How to develop a dendrogram?

Step 1 — Assign each item to its own cluster, such that if you hve N number of items, you have N number of clusters

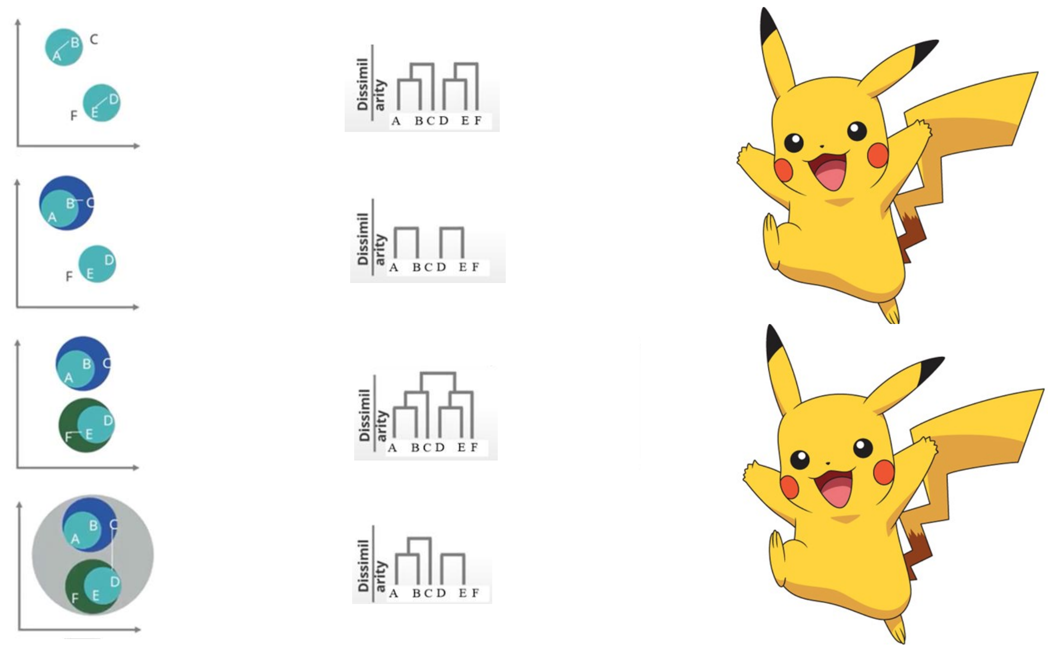
Step 2 — Find the closet (most similar) pair of clusters and merge them into a single cluster. Now you have one cluster less.

Step 3 — Compute distances (similarities) between the new cluster and every old cluster.

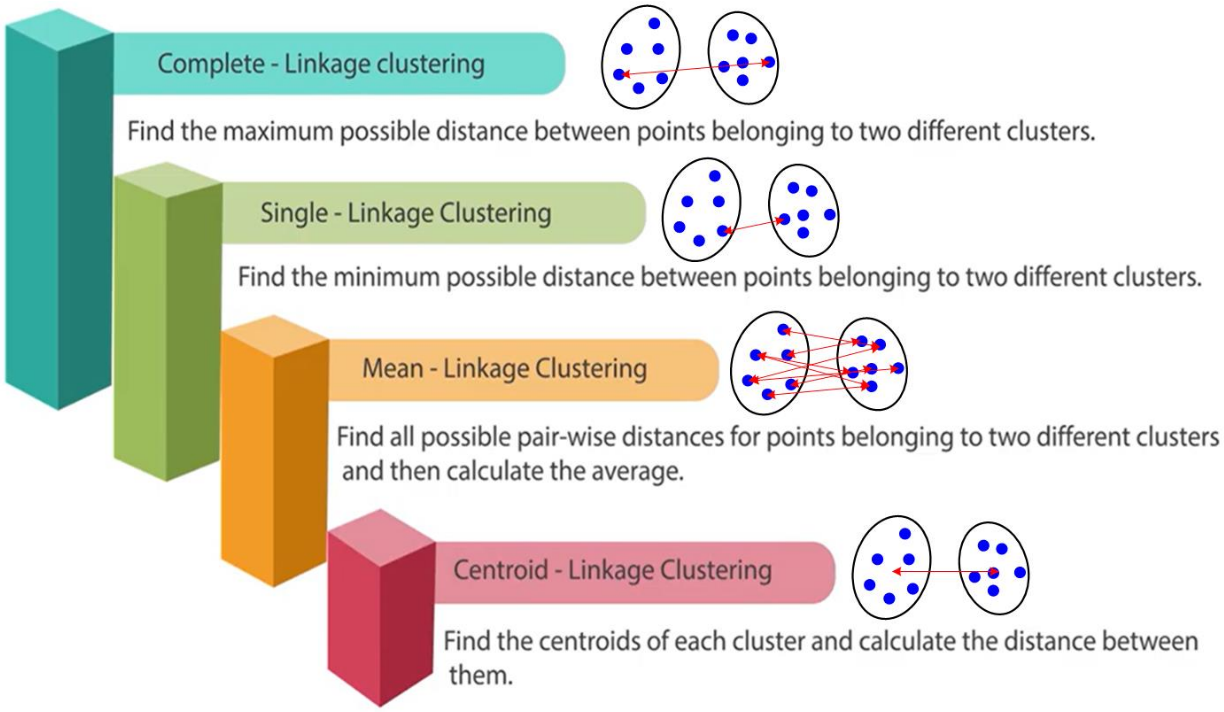
Step 4 — Repeat steps two and three until all items are clustered into a single cluster of size N.

Quiz 2.1:

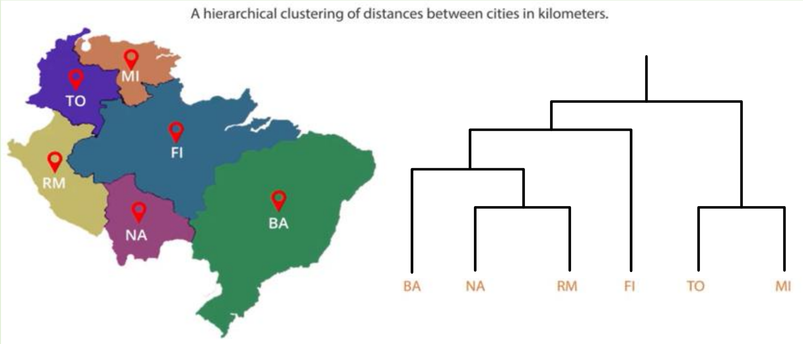
Match the steps for a dendrogram development:



Distance measurements:



Example: Clustering of cities



Step 1:

	BA	FI	MI	NA	RM	TO
BA	0	662	877	255	412	996
FI	662	0	295	468	268	400
MI	877	295	0	754	564	138
NA	255	468	754	0	219	869
RM	412	268	564	219	0	669
TO	996	400	138	869	669	0

MI/TO

MI<TO

	BA	FI	MI/TO	NA	RM
BA	0	662	877	255	412
FI	662	0	295	468	268
MI/TO	877	295	0	754	564
NA	255	468	754	0	219
RM	412	268	564	219	0

Step 2:

	BA	FI	MI/TO	NA	RM
BA	0	662	877	255	412
FI	662	0	295	468	268
MI/TO	877	295	0	754	564
NA	255	468	754	0	219
RM	412	268	564	219	0

RM/NA

Minimum

	BA	FI	MI/TO	NA/RM
BA	0	662	877	255
FI	662	0	295	268
MI/TO	877	295	0	564
NA/RM	255	468	564	0

Step 3:

	BA	FI	MI/TO	NA/RM
BA	0	662	877	255
FI	662	0	295	268
MI/TO	877	295	0	564
NA/RM	255	268	564	0

BA/(RM/NA)

Minimum

	BA/(NA/RM)	FI	MI/TO
BA/(NA/RM)	0	268	564
FI	268	0	295
MI/TO	564	295	0

Step 4:

	BA/(NA/RM)	FI	MI/TO
BA/(NA/RM)	0	268	564
FI	268	0	295
MI/TO	564	295	0

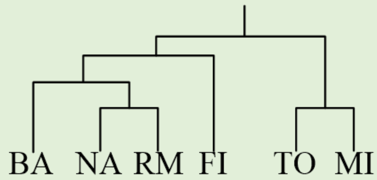
Minimum

FI/[BA/(NA/RM)]

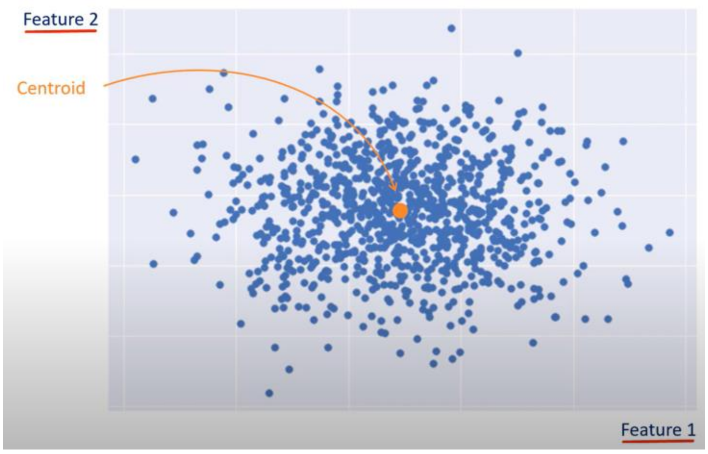
	BA/(NA/RM)/FI	(MI/TO)
BA/(NA/RM)/FI	0	295
(MI/TO)	295	0

Step 5:

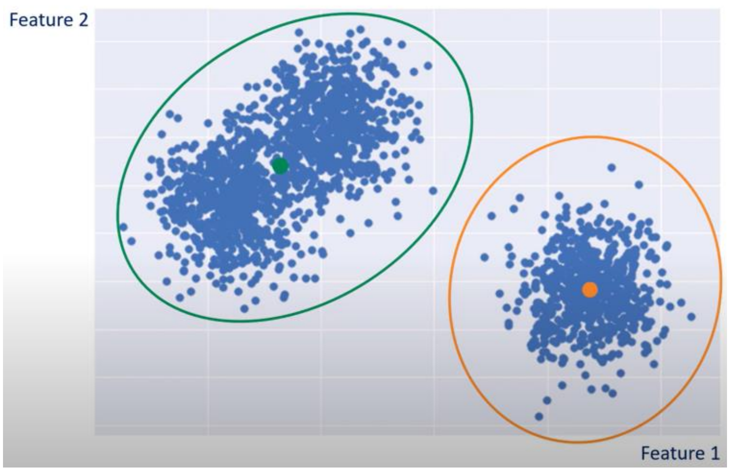
	BA/(NA/RM)/FI	(MI/TO)
BA/(NA/RM)/FI	0	295
(MI/TO)	295	0



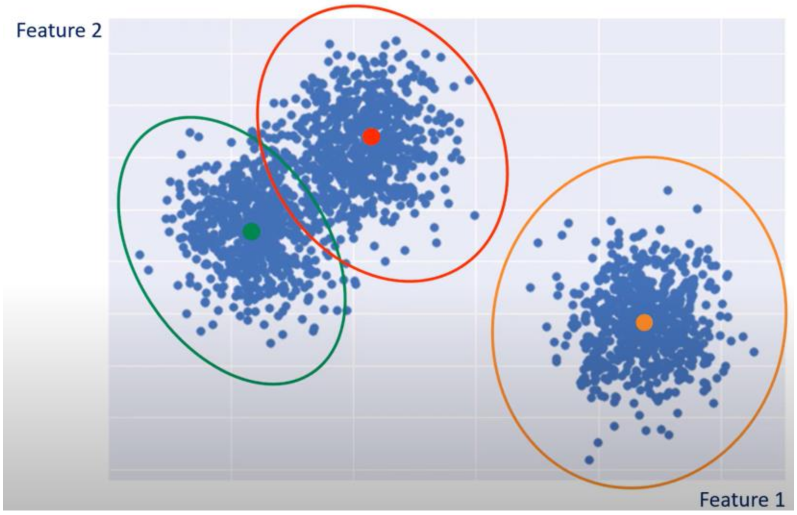
3.2 Centroid-based clustering – K-means



We want relevant data crowded around a centroid as one cluster.



We may have more clusters

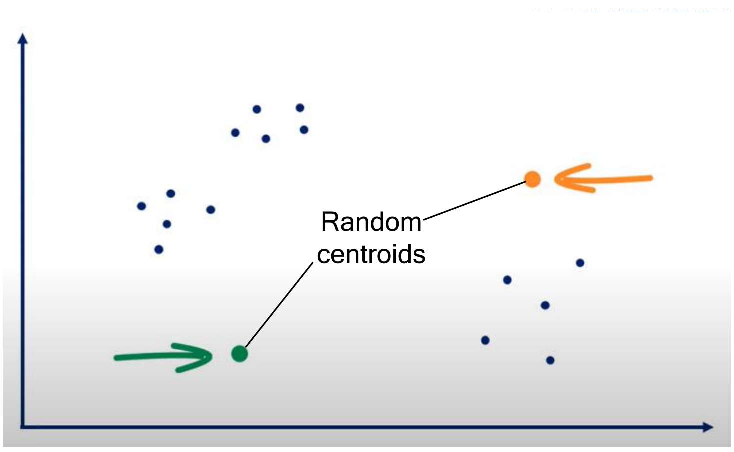


How can we generate these clusters in practice? K-means **[4]** is the most popular way to do this.

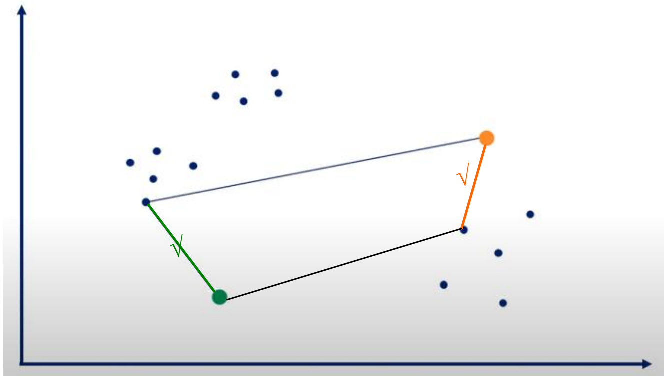
Let’s simplify the graph into 15 points below.

1. Choose the number of clusters (K, i.e. K=2)

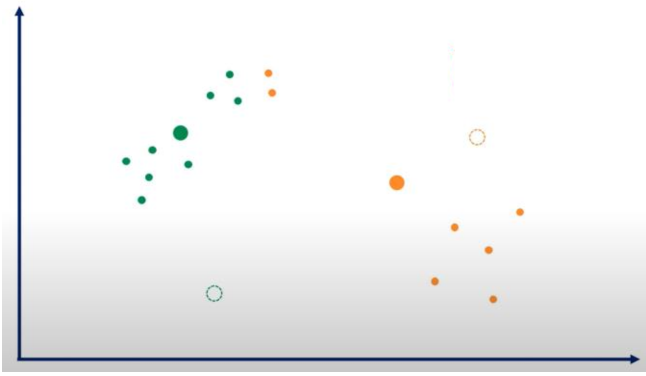
2. Specify the cluster seeds

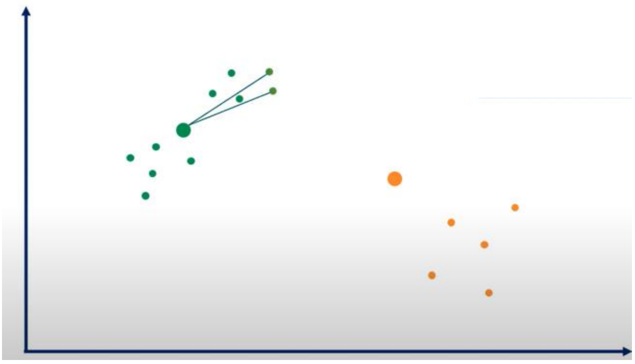


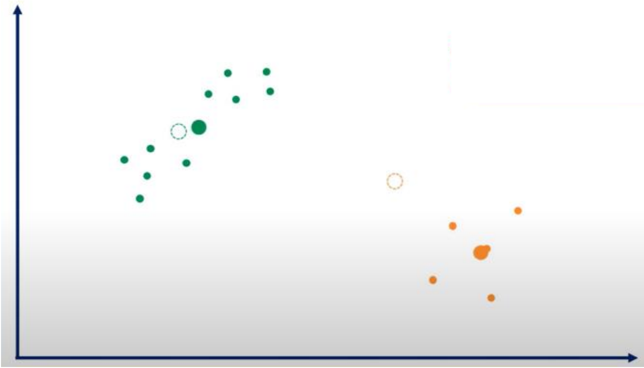
3. Assign each point to a centroid



4. Adjust the centroid

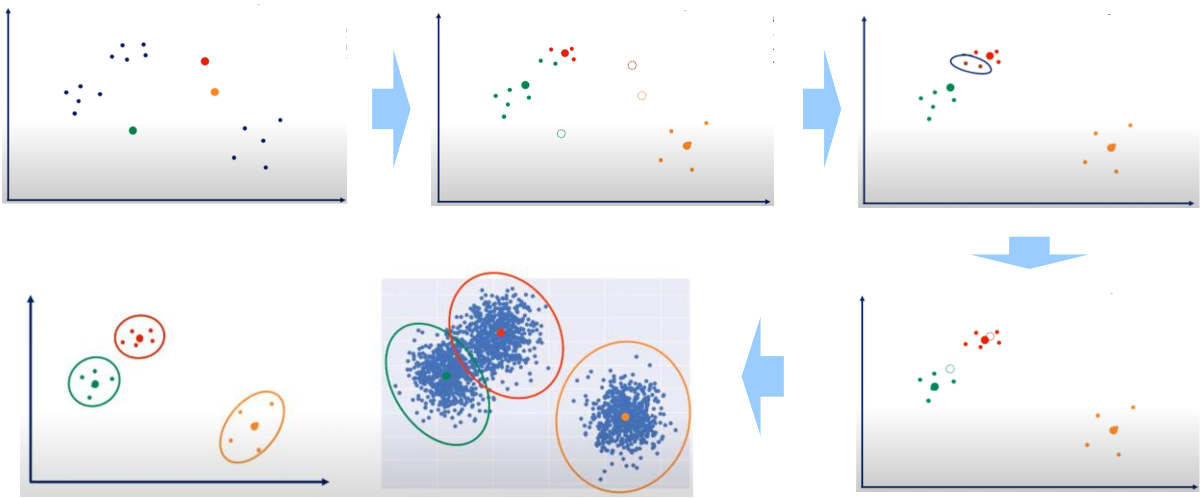






Now all green points are close to green centroid and all orange points are close to orange centroid. The calculation terminates.

For K=3, we follow the exact same process to achieve the cluster results:

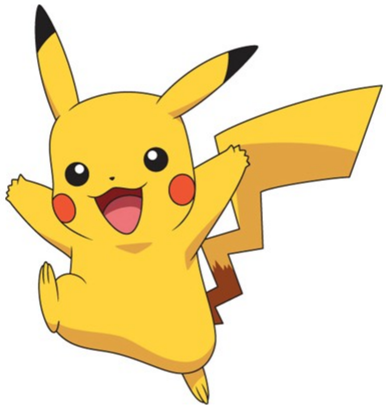


In above examples, the Euclidean distance is applied. The Euclidean distance between two points $\mathbf{p}=(p_1,\ldots,p_n),\mathbf{q}=(q_1,\ldots,q_n)$ is

$$D=\sqrt{(p_1-q_1)^2+\cdots+(p_n-q_n)^2}$$

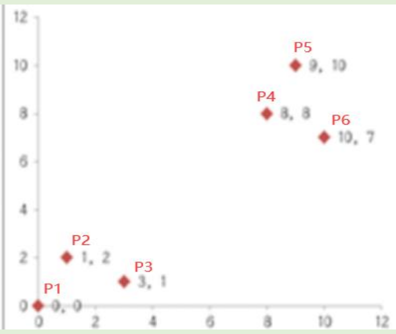
Quiz 3.1:

Find the distance between $\mathbf{p}=(1,2),\mathbf{q}=(3,4)$



Example: K-means

	X	Y
P1	0	0
P2	1	2
P3	3	1
P4	8	8
P5	9	10
P6	10	7



Find 2 groups by using K-means clustering.

Step 1: Select the initial centroids: P1, P2

Step 2: Find the Euclidean distances from P3-P6 to P1 and P2:

	P1	P2
P3	3.16	2.24
P4	11.3	9.22
P5	13.5	11.3
P6	12.2	10.3

$$D = \sqrt{(3-0)^2 + (1-0)^2} = \sqrt{10} = 3.16$$

Then the clustering results: $\begin{cases} K = 1: P1 \\ K = 2: P2, P3, P4, P5, P6 \end{cases}$

Step 3: Update centroids and recalculate the Euclidean distances:

$$\begin{cases} K = 1: P1 \rightarrow (0,0) \\ K = 2: Pc \rightarrow (\frac{1+3+8+9+10}{5}, \frac{2+1+8+10+7}{5}) = (6.2, 5.6) \end{cases}$$

	P1	Pc
P2	2.24	6.3246
P3	3.16	5.6036
P4	11.3	3
P5	13.5	5.2154
P6	12.2	4.0497

Then the clustering results: $\begin{cases} K = 1: P1, P2, P3 \\ K = 2: P4, P5, P6 \end{cases}$

Step 4: Update centroids and recalculate the Euclidean distances:

$$\begin{cases} K = 1: Pc1 \rightarrow (1.33, 1) \\ K = 2: Pc2 \rightarrow (9, 8.33) \end{cases}$$

	Pc1	Pc2
P1	1.4	12
P2	0.6	10
P3	1.4	9.5
P4	47	1.1
P5	70	1.7
P6	56	1.7

$$\Rightarrow \begin{cases} K = 1: P1, P2, P3 \\ K = 2: P4, P5, P6 \end{cases}$$

Then the cluster is convergent.

4 Further Readings

[1] CNN.

https://en.wikipedia.org/wiki/Convolutional_neural_network

[2] Convolution

https://www.inf.ed.ac.uk/teaching/courses/cfcs1/lectures/cfcs_l15.pdf

[3] Hierarchical clustering.

https://en.wikipedia.org/wiki/Hierarchical_clustering

[4] K-means clustering

https://en.wikipedia.org/wiki/K-means_clustering