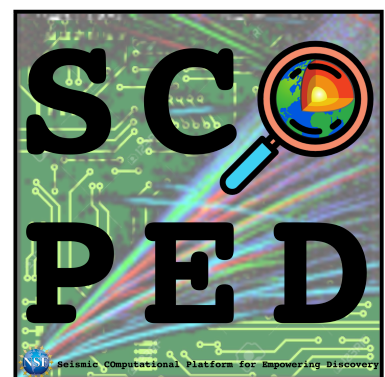




Lamont-Doherty Earth Observatory
COLUMBIA UNIVERSITY | EARTH INSTITUTE



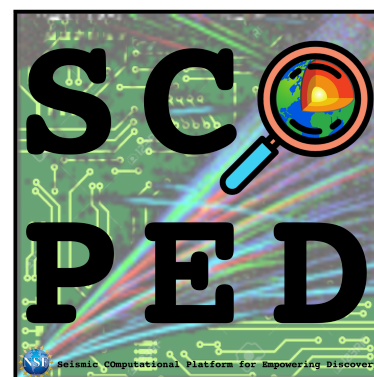
Unsupervised Machine Learning for Seismology Using AWS

SCOPED University of Washington, Seattle 2024

Theresa Sawi, Nate Groebner, Felix Waldhauser, Kaiwen Wang, Yen-Joe Tan

Thursday 5/23/2024

1:00 - 1:45	Lecture on unsupervised machine learning in seismology
1:45 - 2:00	Q&A
2:00 - 2:10	Break
2:10 - 3:00	Tutorial on SpecUFEEx/2016 Amatrice sequence



Lamont-Doherty Earth Observatory
COLUMBIA UNIVERSITY | EARTH INSTITUTE



Unsupervised Machine Learning for Seismology Using AWS

SCOPED University of Washington, Seattle 2024

Theresa Sawi, Nate Groebner, Felix Waldhauser, Kaiwen Wang, Yen-Joe Tan

Last updated 05/23/2024

Outline

- **Introduction to unsupervised machine learning (UML)**
 - feature extraction & dimensionality reduction
 - clustering
- **SpecUFEx tutorial: Amatrice 2016**

Outline

- **Introduction to unsupervised machine learning (UML)**
 - feature extraction & dimensionality reduction
 - clustering
- **SpecUFEx tutorial: Amatrice 2016**

Supervised Machine Learning

Labeled (training) dataset

Cats



Dogs



Supervised Machine Learning

Labeled (training) dataset

Cats



Dogs

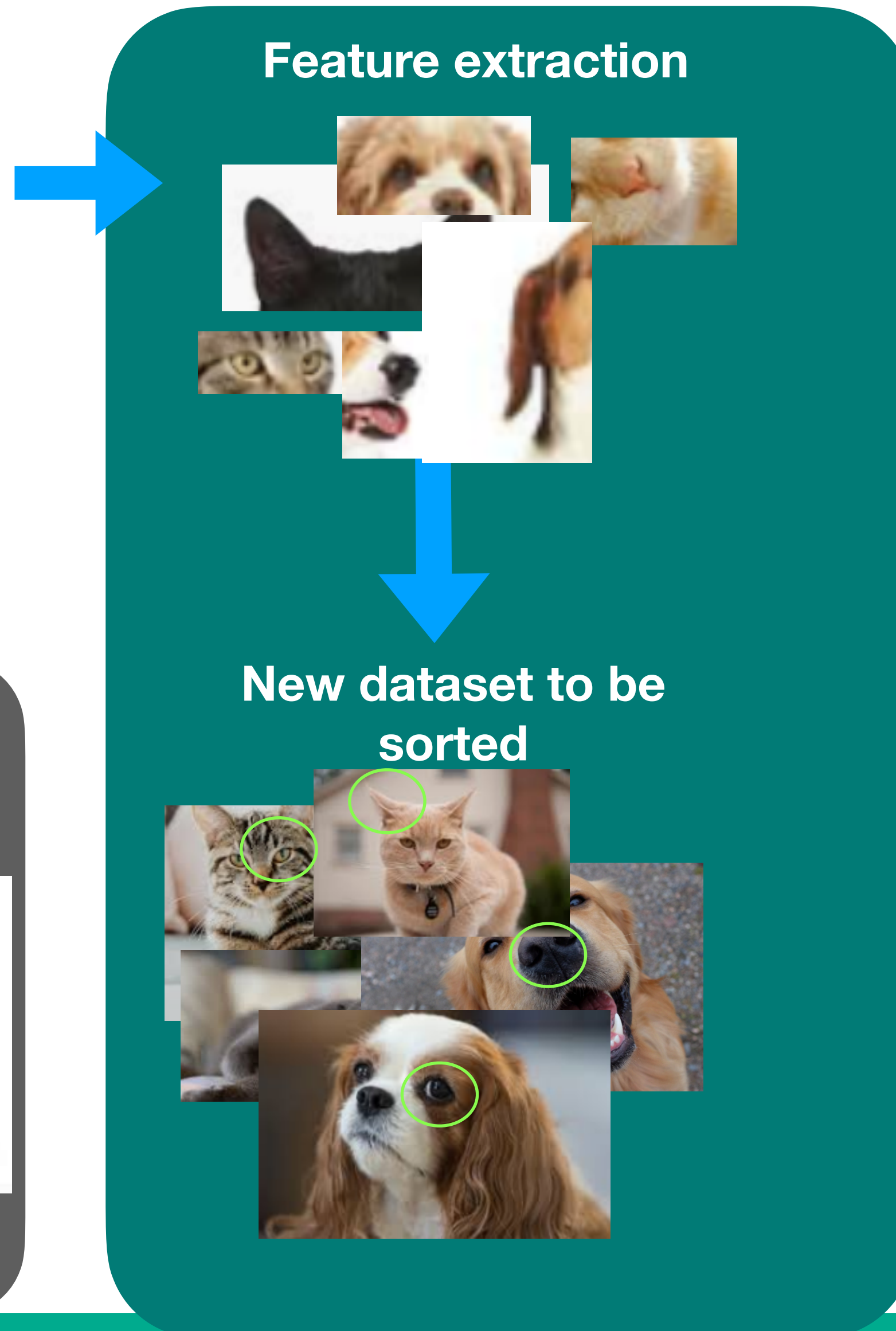


New dataset to be sorted



Supervised Machine Learning

Labeled (training) dataset

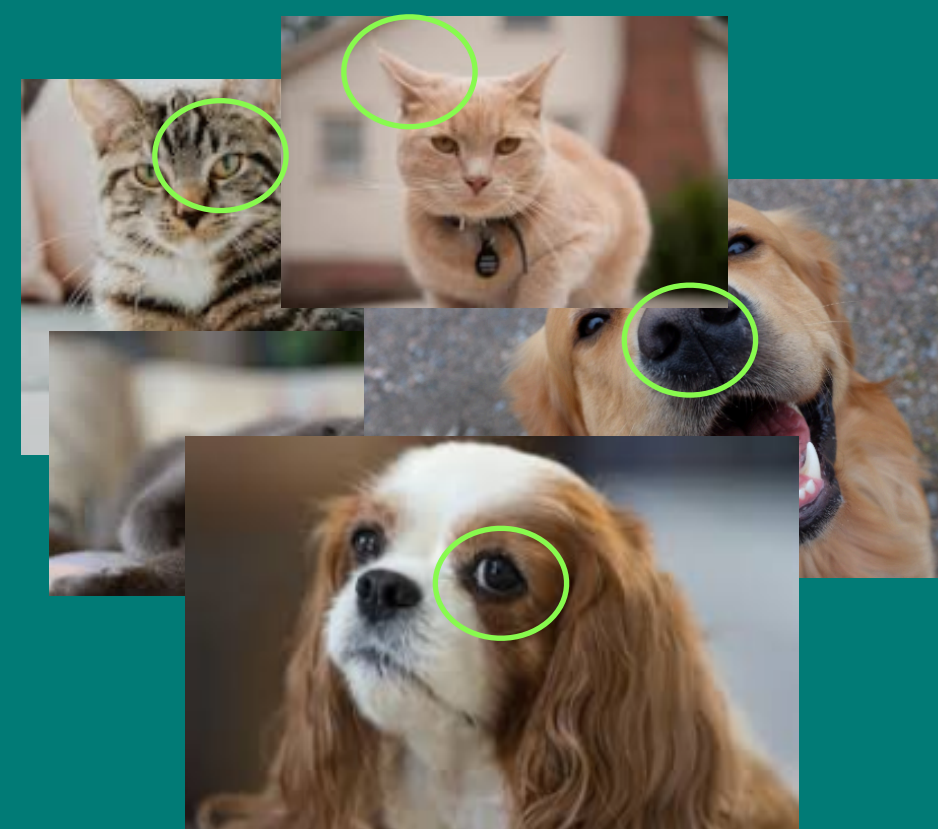


Supervised Machine Learning

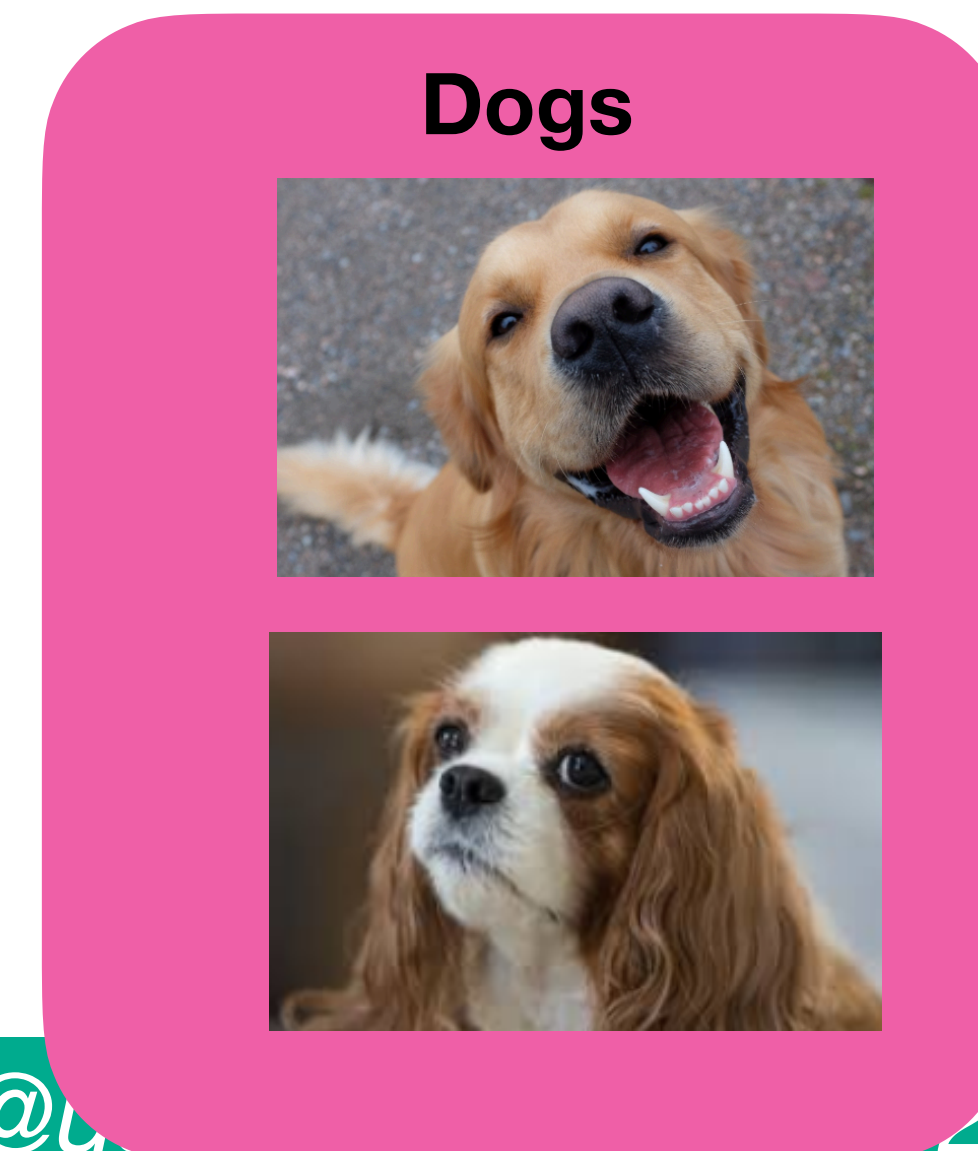
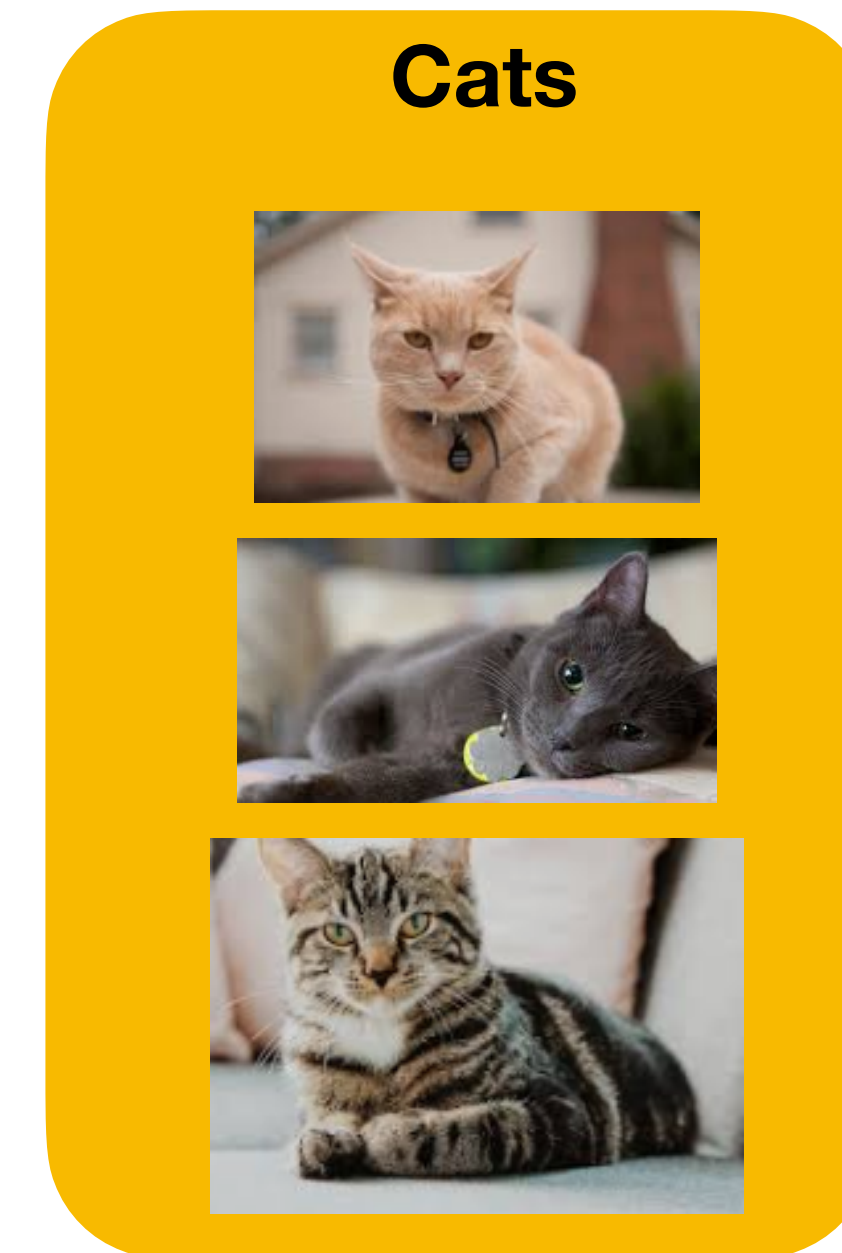
Labeled (training) dataset



Feature extraction



Classification



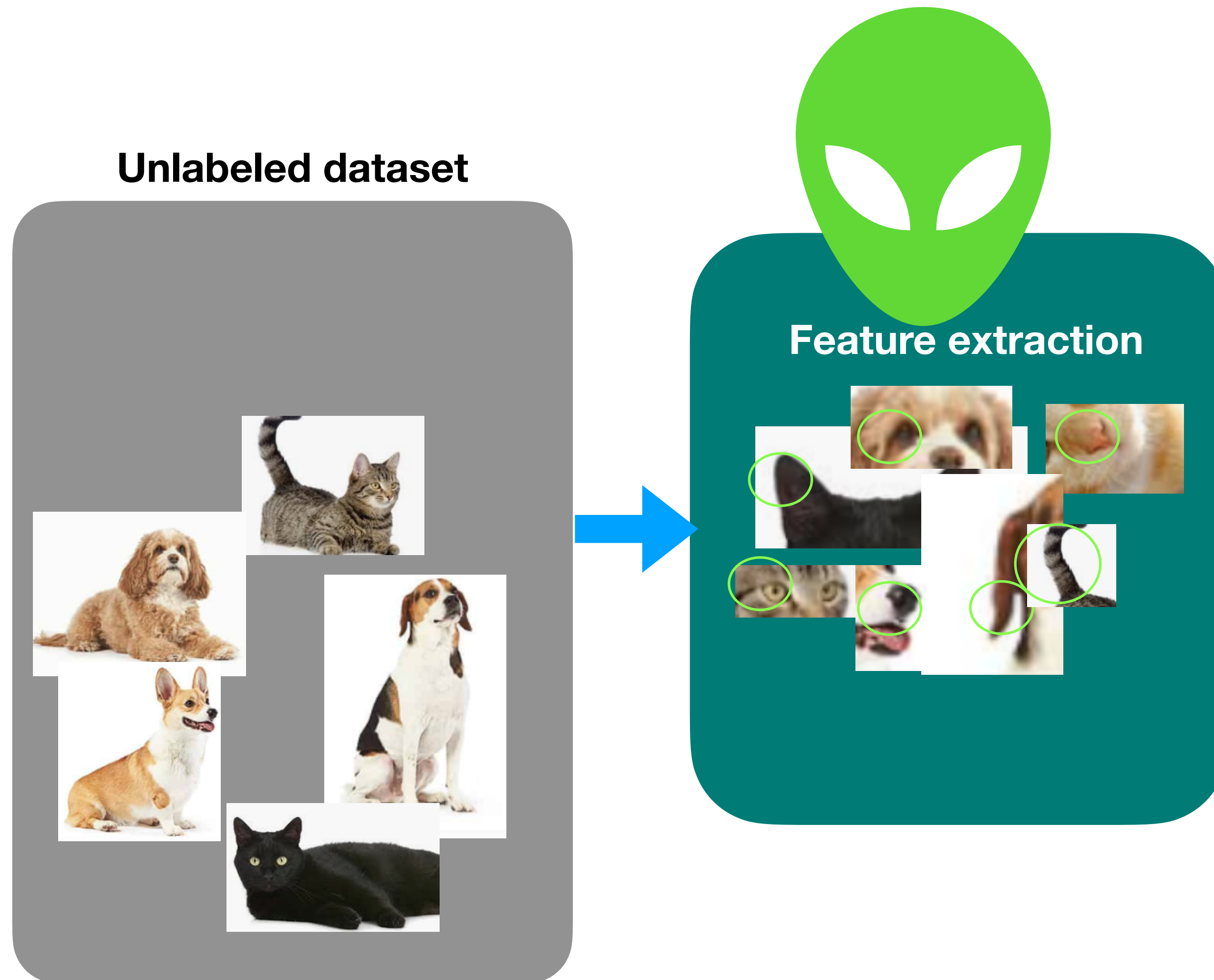
Unsupervised Machine Learning



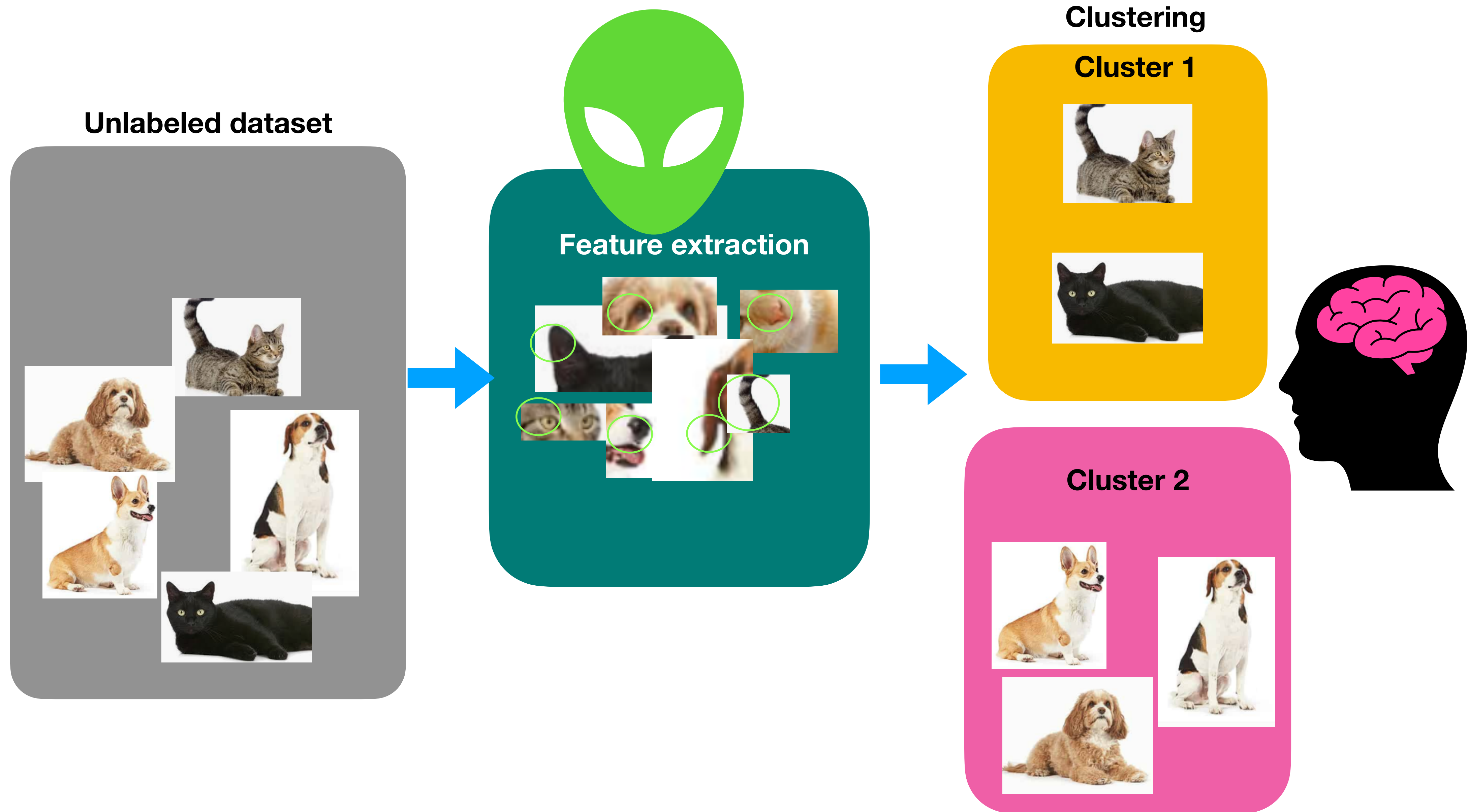
Unlabeled dataset



Unsupervised Machine Learning



Unsupervised Machine Learning



Machine Learning

Supervised

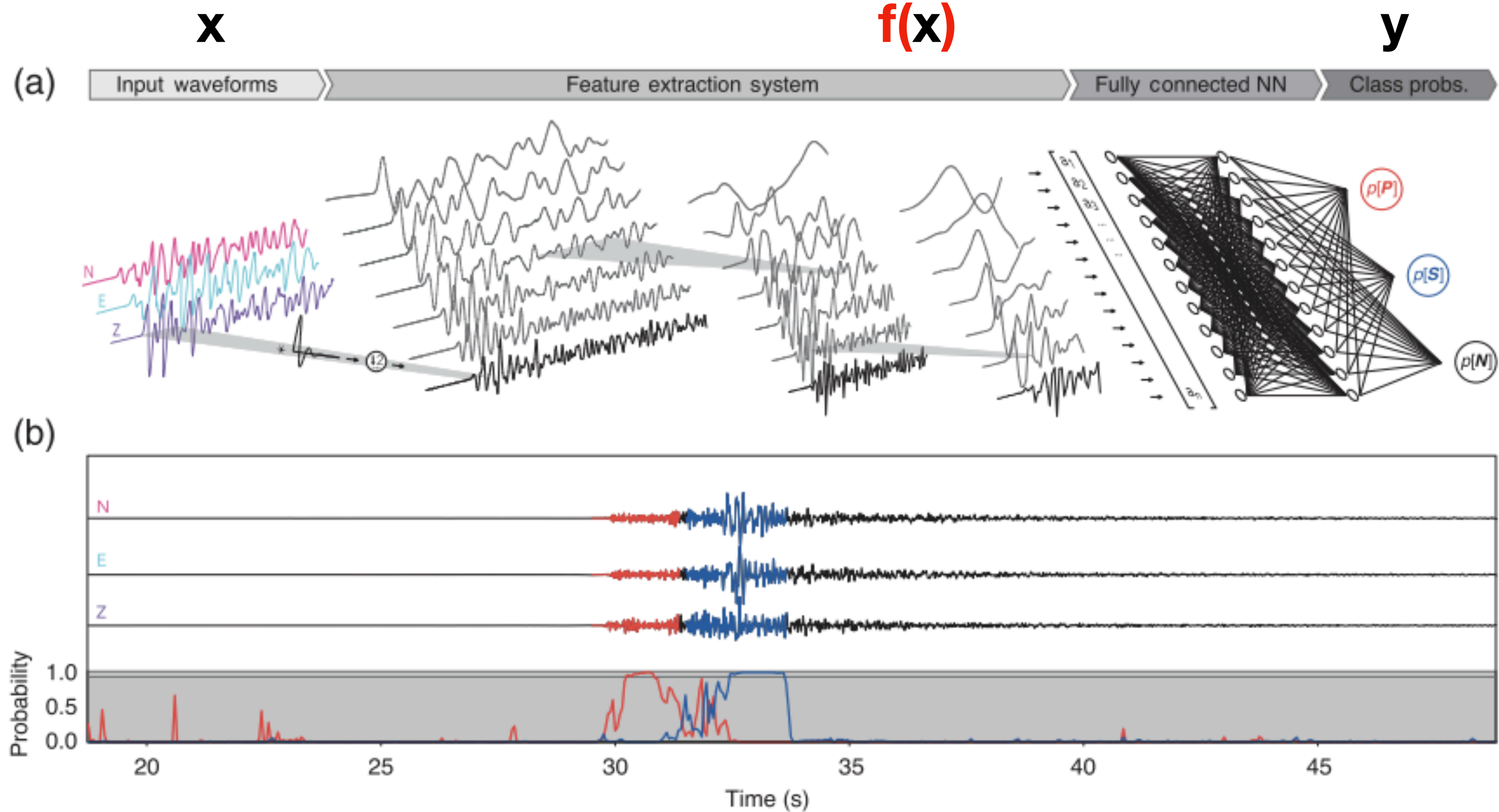
Unsupervised

Train a **function** that best
approximates mapping
between input data (x) and
labels (y)

$$\mathbf{f(x) = y}$$

Supervised Machine Learning

$f(x)$ trained on millions of P- and S-wave arrivals



Ross et al., 2018, BSSA

Machine Learning

Supervised

Train a **function** that best approximates mapping between input data (x) and labels (y)

$$\mathbf{f(x) = y}$$

Unsupervised

Learn a **function** that *infers* natural structure within the data

$$\mathbf{f(x) = y}$$

Machine Learning

Supervised

Train a **function** that best approximates mapping between input data (x) and labels (y)

$$f(x) = y$$

- Needs labeled data
- Only finds patterns that it has learned
- Routine work

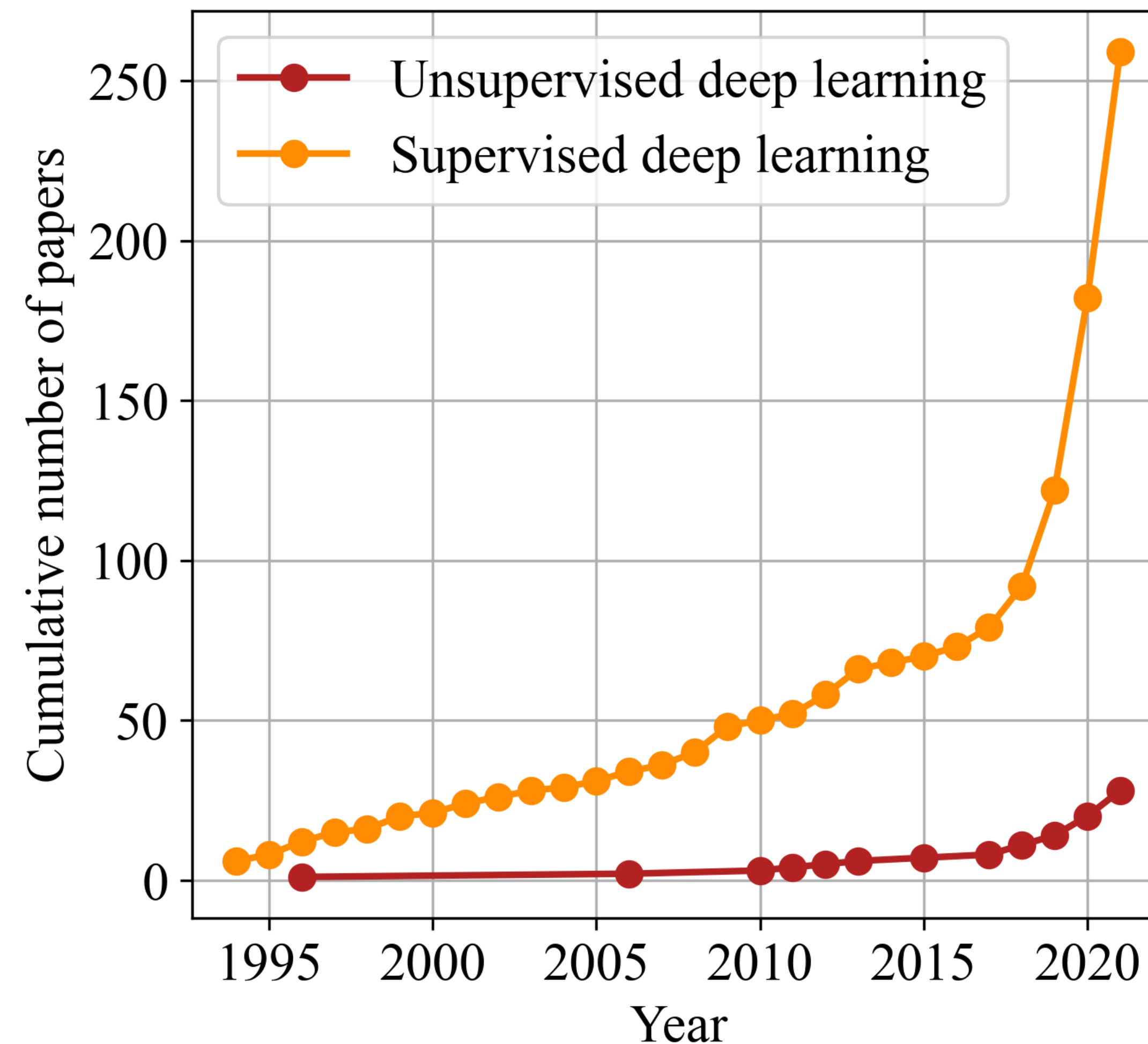
Unsupervised

Learn a **function** that *infers* natural structure within the data

$$f(x) = y$$

- Unlabeled data
- Infers patterns in data
- Exploratory work

Unsupervised machine learning (UML) in seismology



Sawi Dissertation, 2023. Adapted from Mousavi & Beroza, 2022; Science

ML Algorithm Examples

<u>Unsupervised ML</u>	<u>Supervised ML</u>
Dimensionality reduction: Principal component analysis (PCA) Singular value decomposition (SVD) Nonnegative matrix factorization (NMF)*	Regression: Linear Polynomial
Clustering: Kmeans* Hierarchical DBSCAN	Classification: Random forest K-nearest neighbors Support vector machines (SVM)

Outline

- Introduction to unsupervised machine learning (UML)
 - **feature extraction & dimensionality reduction**
 - clustering
- **SpecUFEx tutorial: Amatrice 2016**

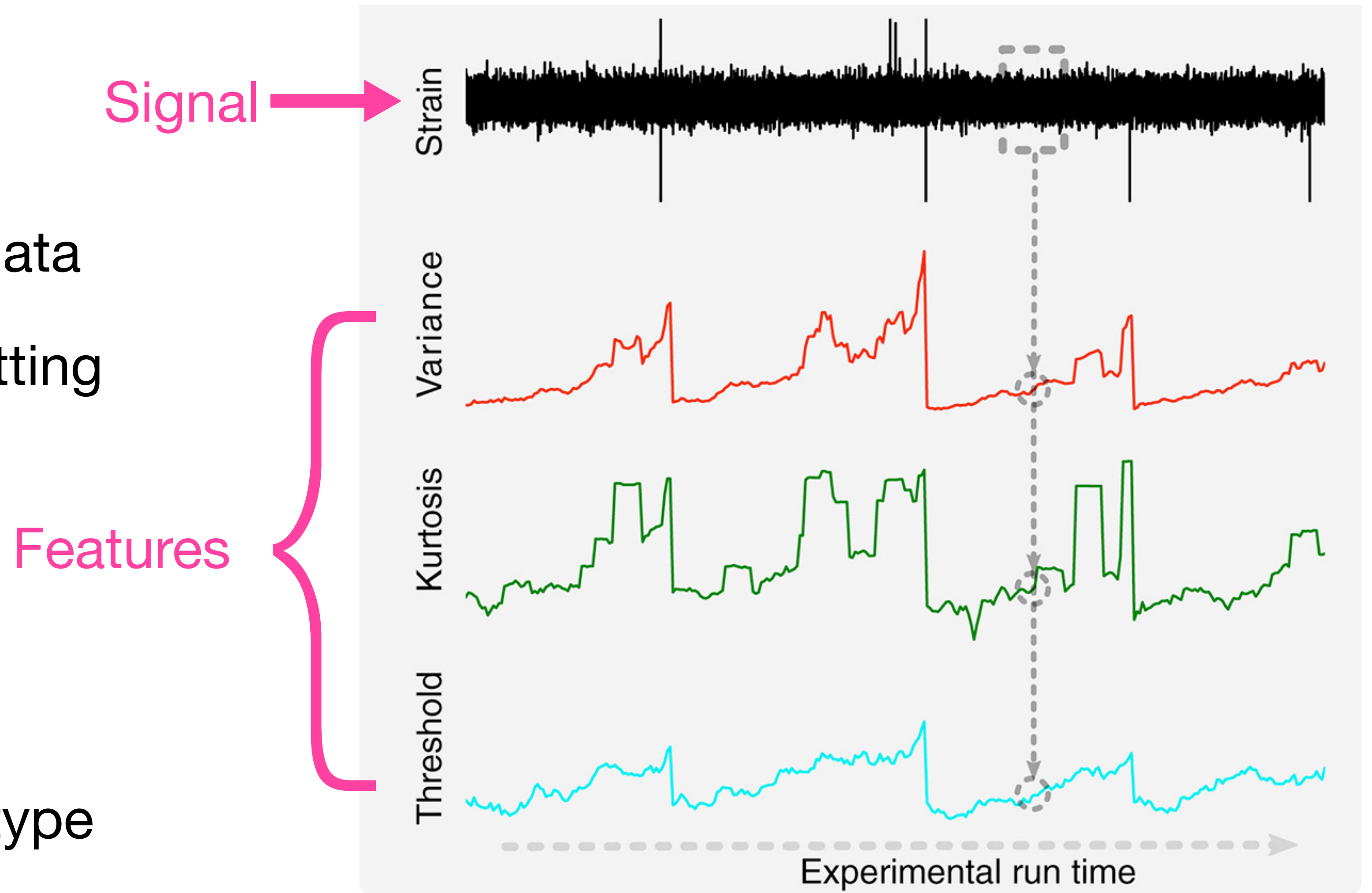
Feature Extraction & Dimensionality Reduction

“Garbage in, garbage out”

Extracting information from data

- Reduce noise and overfitting
- Aid interpretation
- Reduces dimension
- Faster computations

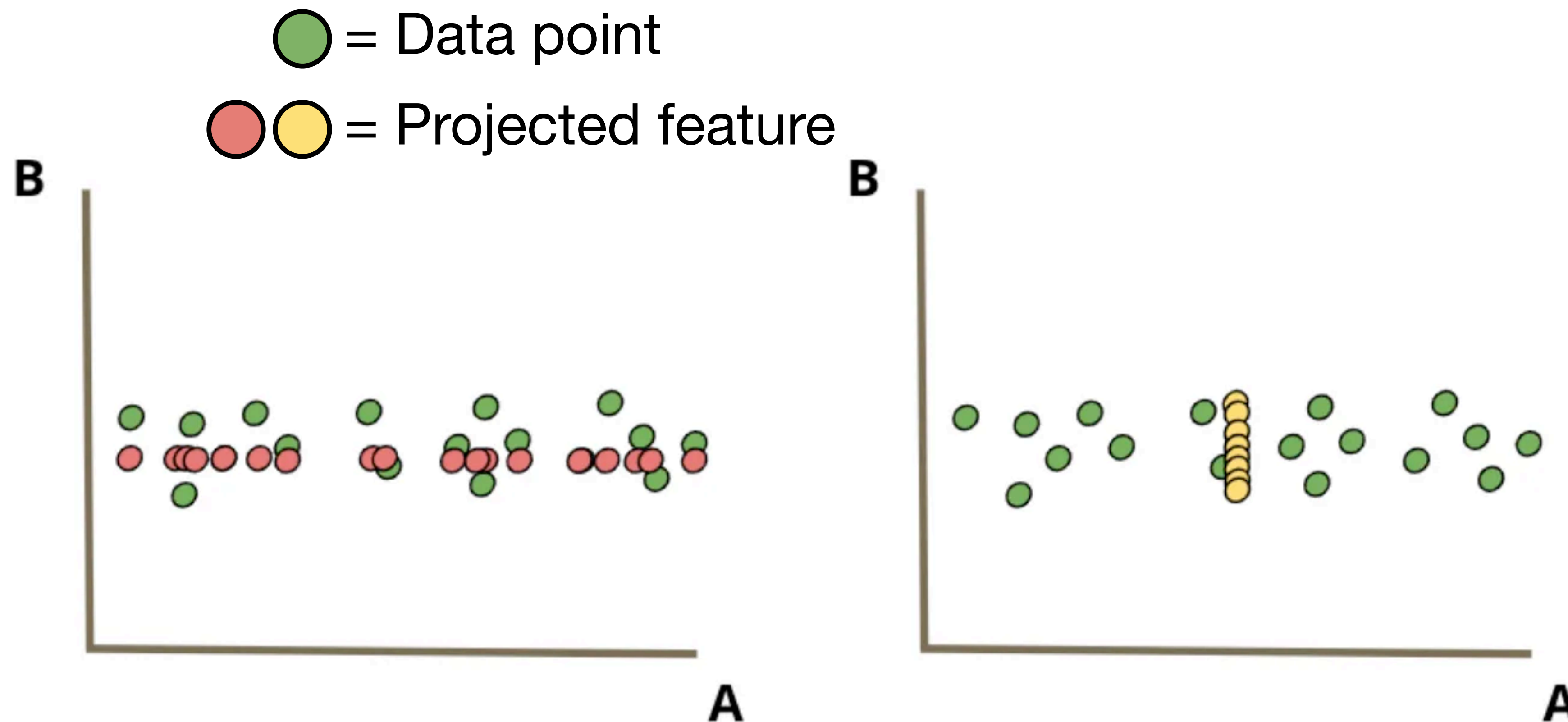
Techniques depend on data type and domain knowledge



Rouet-Leduc et al., (2017) GRL

Feature Extraction & Dimensionality Reduction

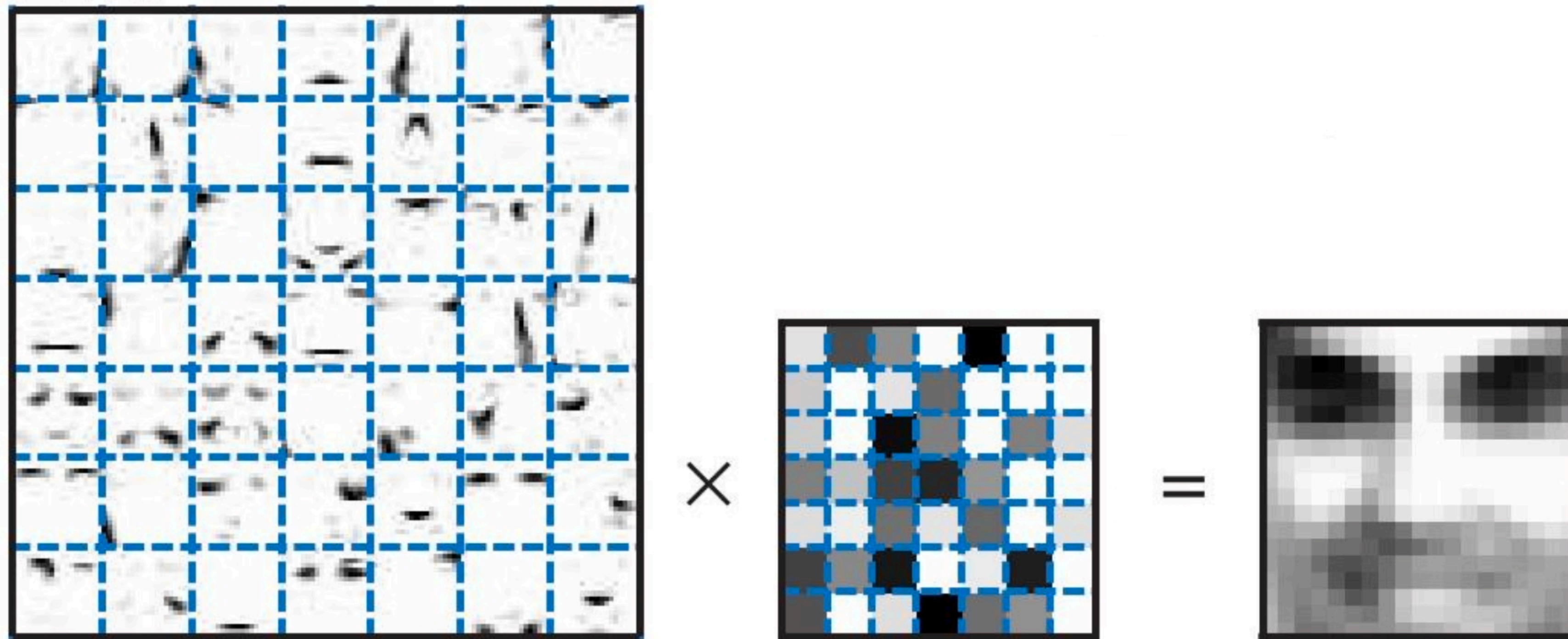
Example: find features with the highest variance / spread
e.g., Singular value decomposition (SVD) and principal component analysis (PCA)



Feature Extraction & Dimensionality Reduction

Example: Non-negative matrix factorization (NMF)

NMF



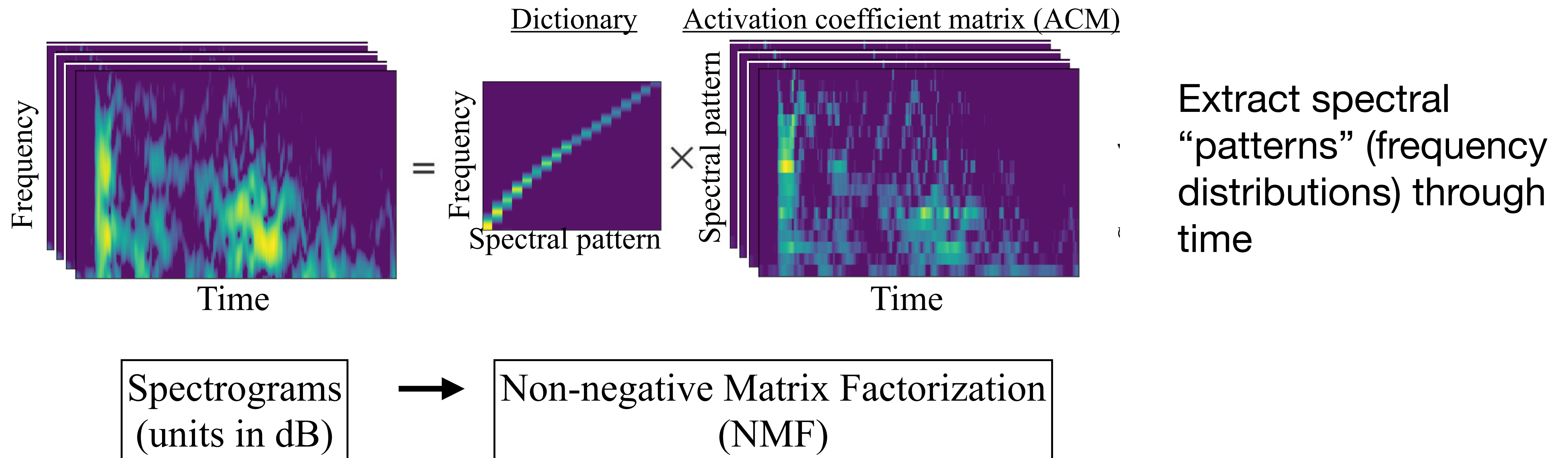
NMF learns a “parts-based” representation. Each column captures something interpretable. This is a result of the nonnegativity constraint.

Lee and H.S. Seung (2001)

Spectral Unsupervised Feature Extraction (SpecUFEx)

Holtzman et al., 2018; Sci Adv

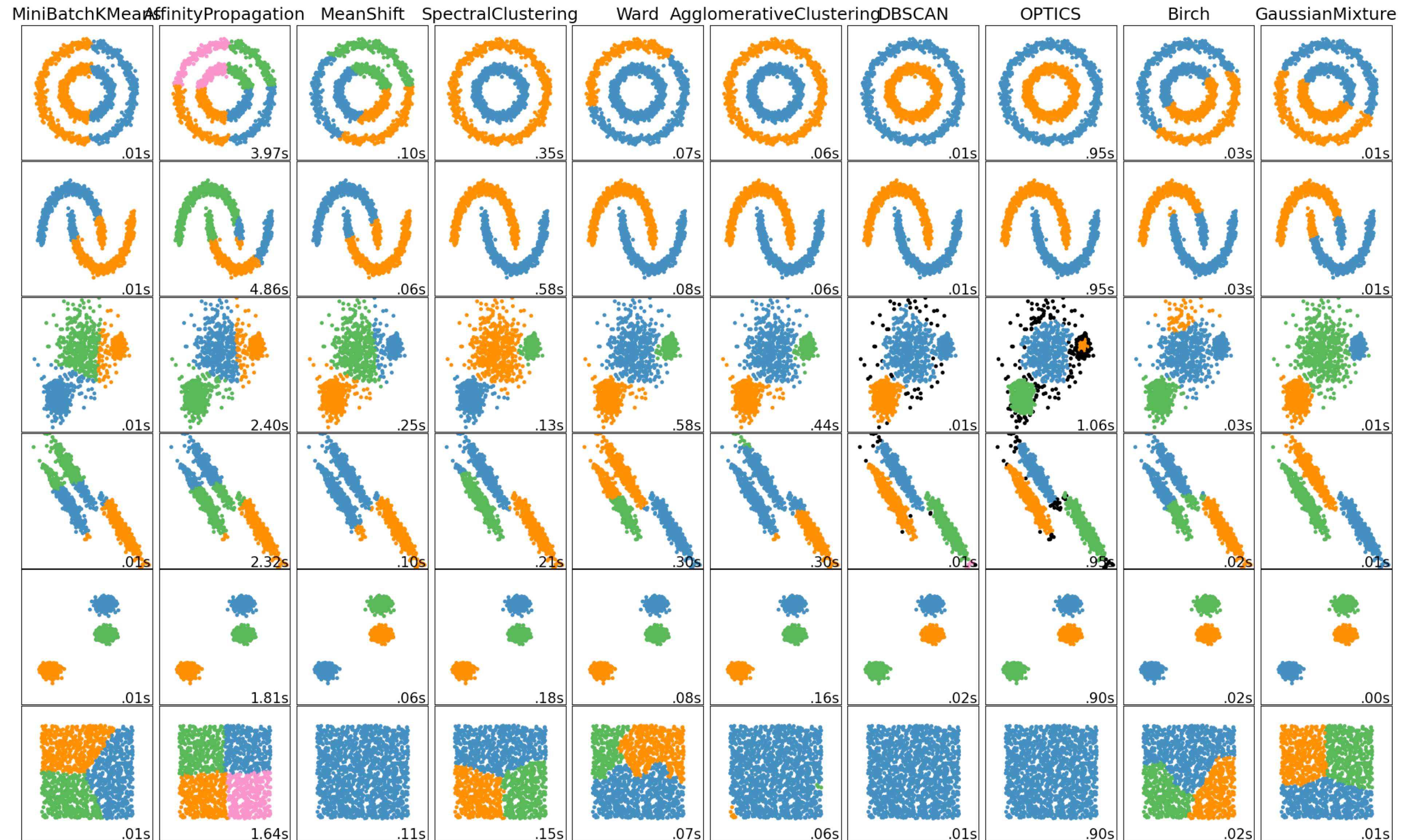
Step 1. Non-Negative Matrix Factorization



Outline

- Introduction to unsupervised machine learning (UML)
 - feature extraction & dimensionality reduction
 - **clustering**
- **SpecUFEx tutorial: Amatrice 2016**

Clustering



Clustering: K-means

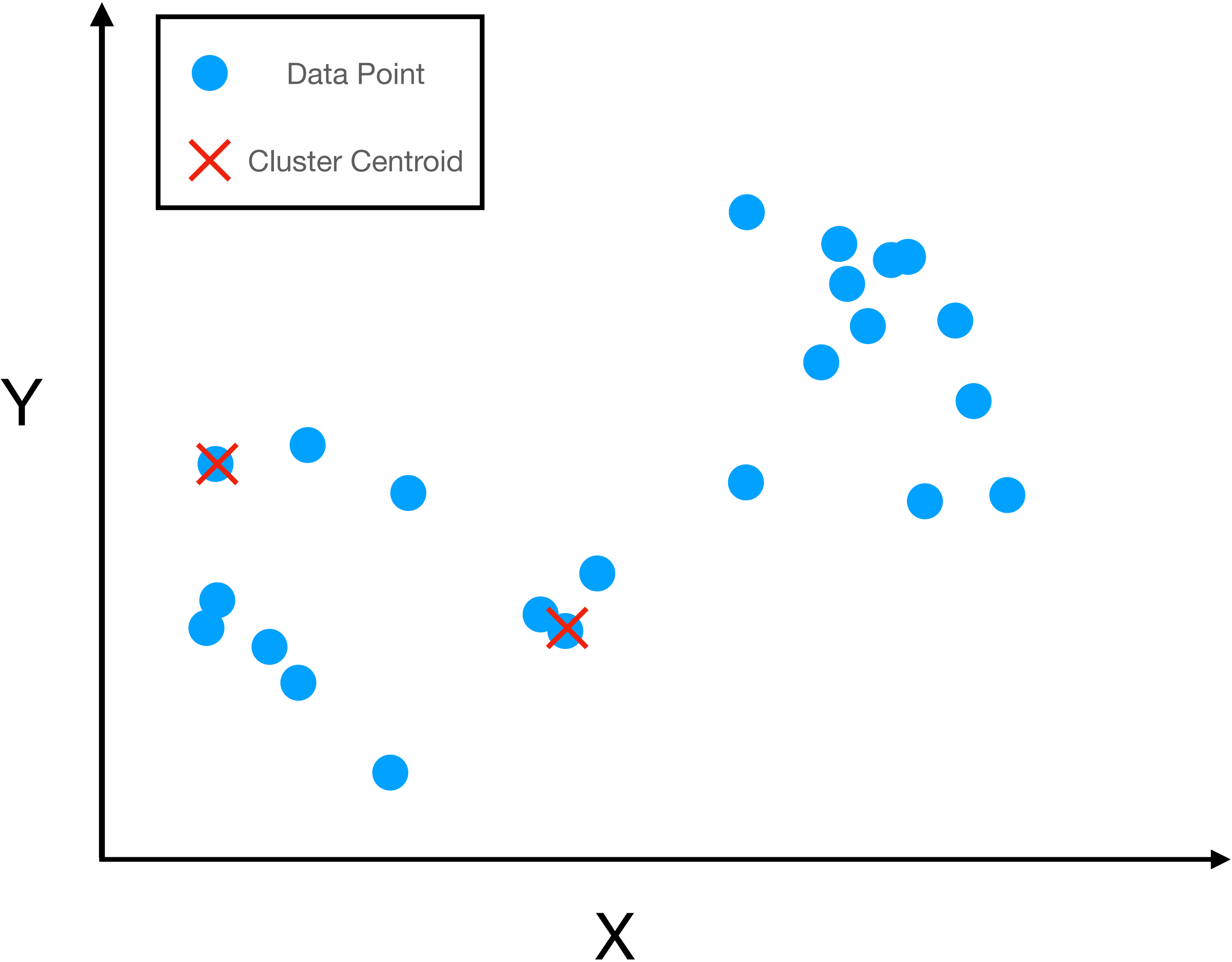
Lloyd, 1982, IEEE



Clustering: K-means

Lloyd, 1982, IEEE

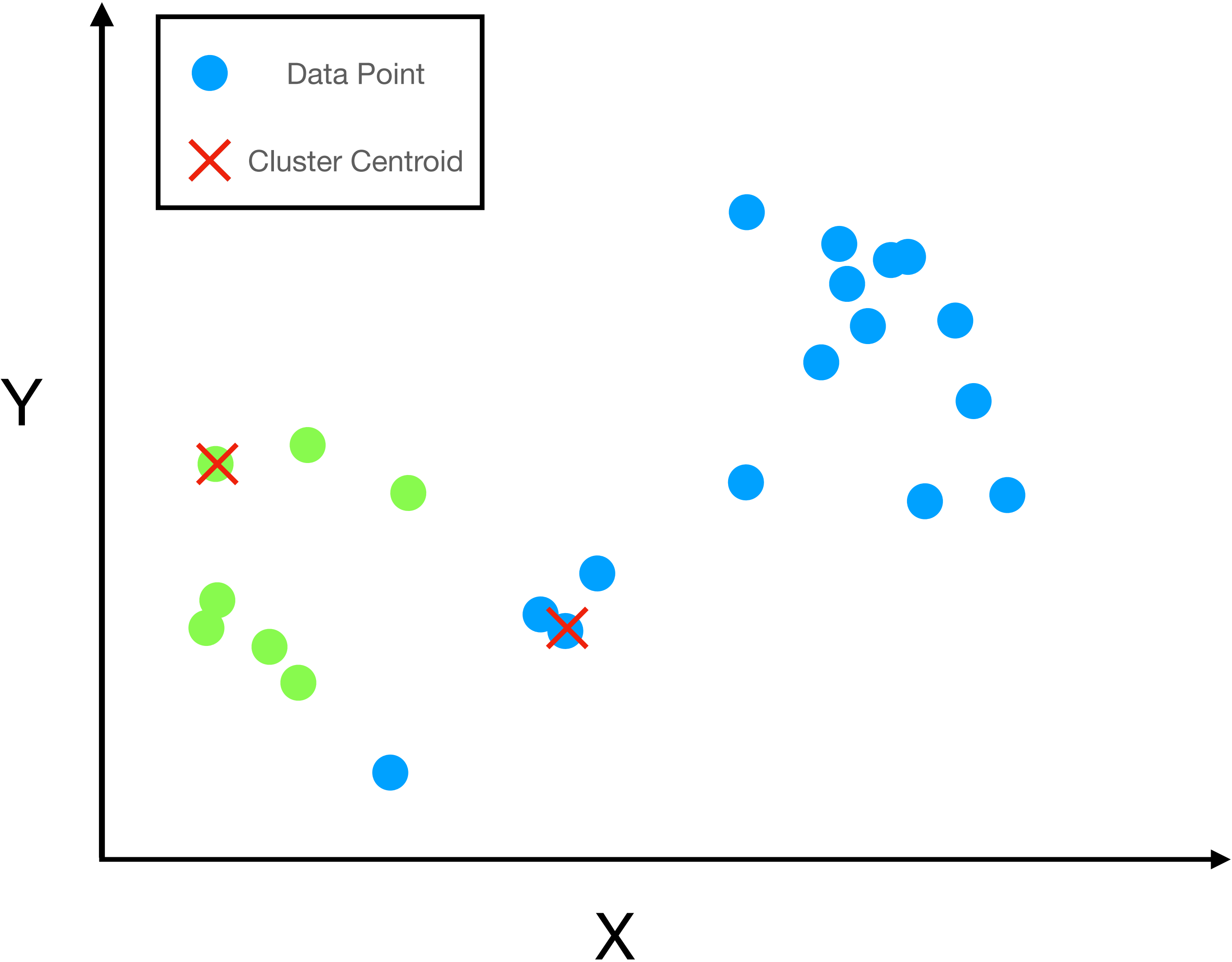
Step 1: Define centroids



Clustering: K-means

Lloyd, 1982, IEEE

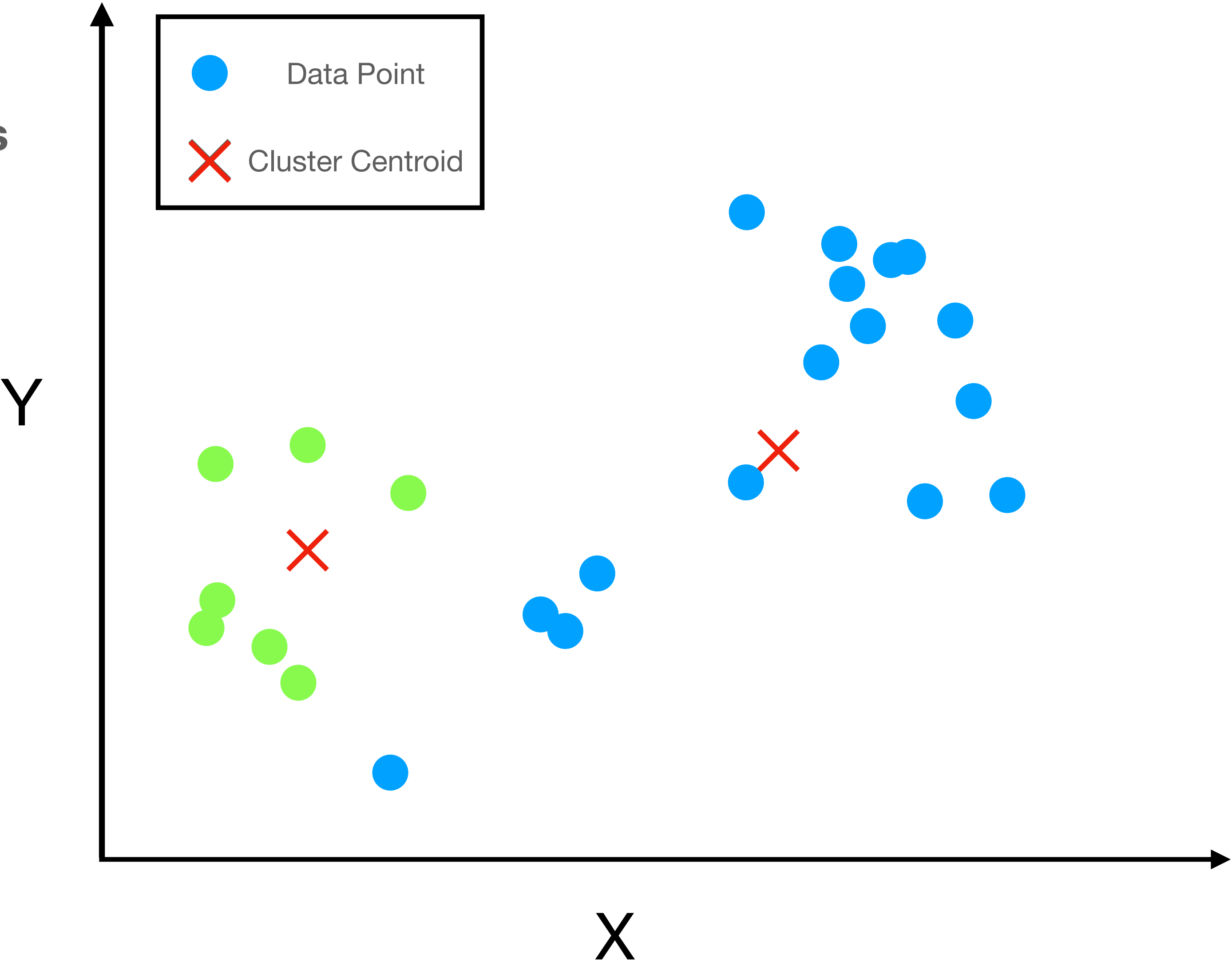
Step 1: Define centroids
Step 2: Cluster data



Clustering: K-means

Lloyd, 1982, IEEE

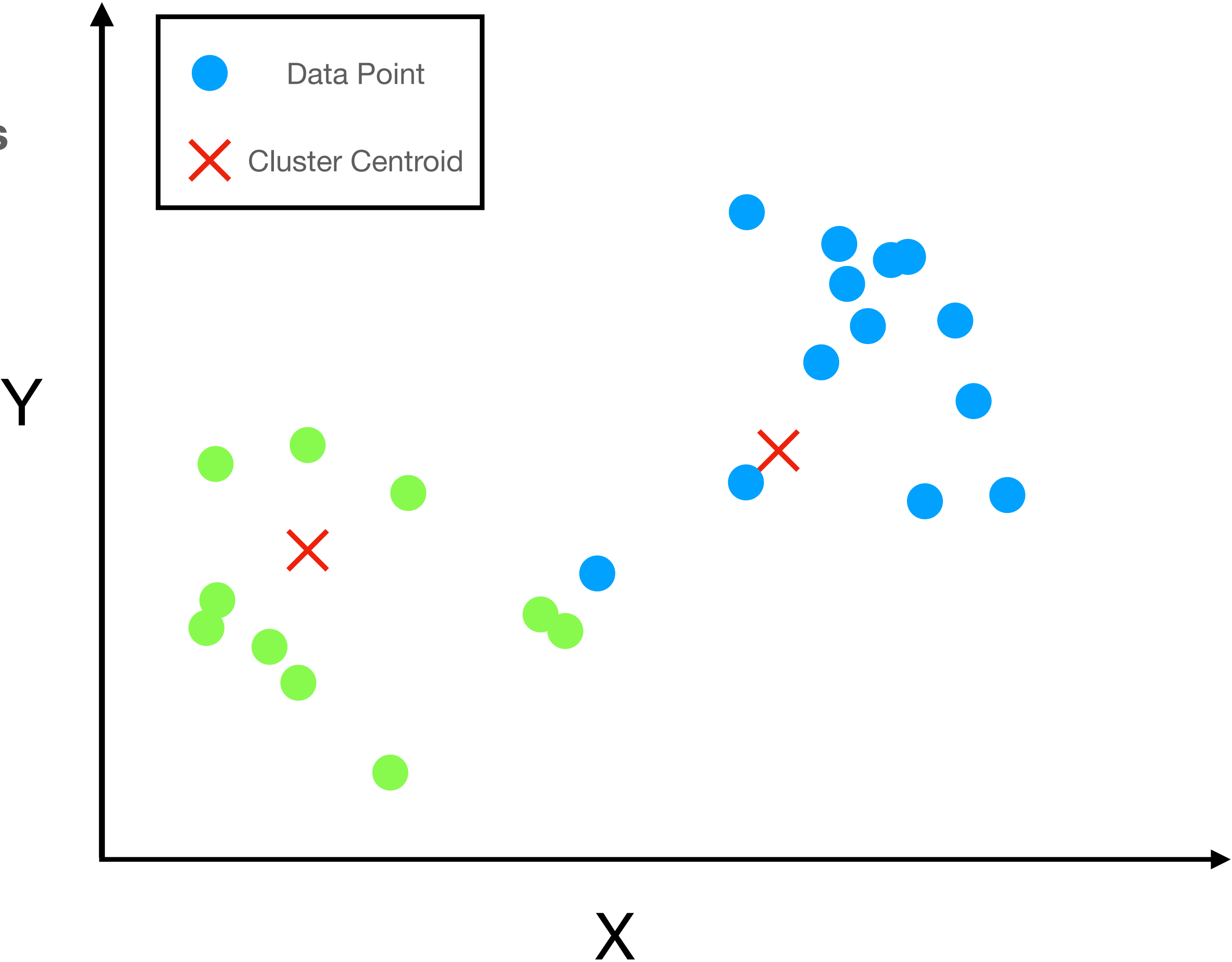
Step 1: Define centroids
Step 2: Cluster data



Clustering: K-means

Lloyd, 1982, IEEE

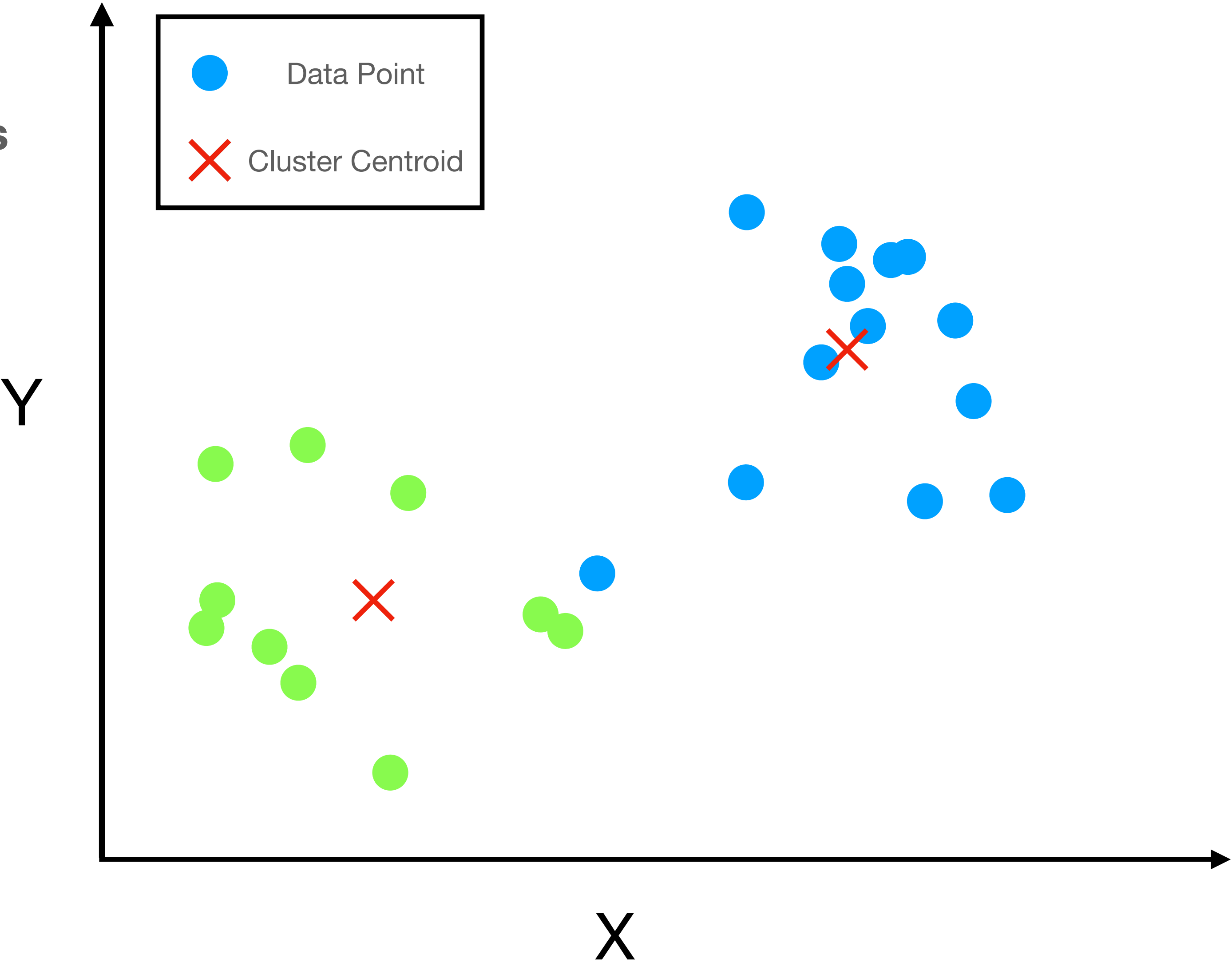
Step 1: Define centroids
Step 2: Cluster data



Clustering: K-means

Lloyd, 1982, IEEE

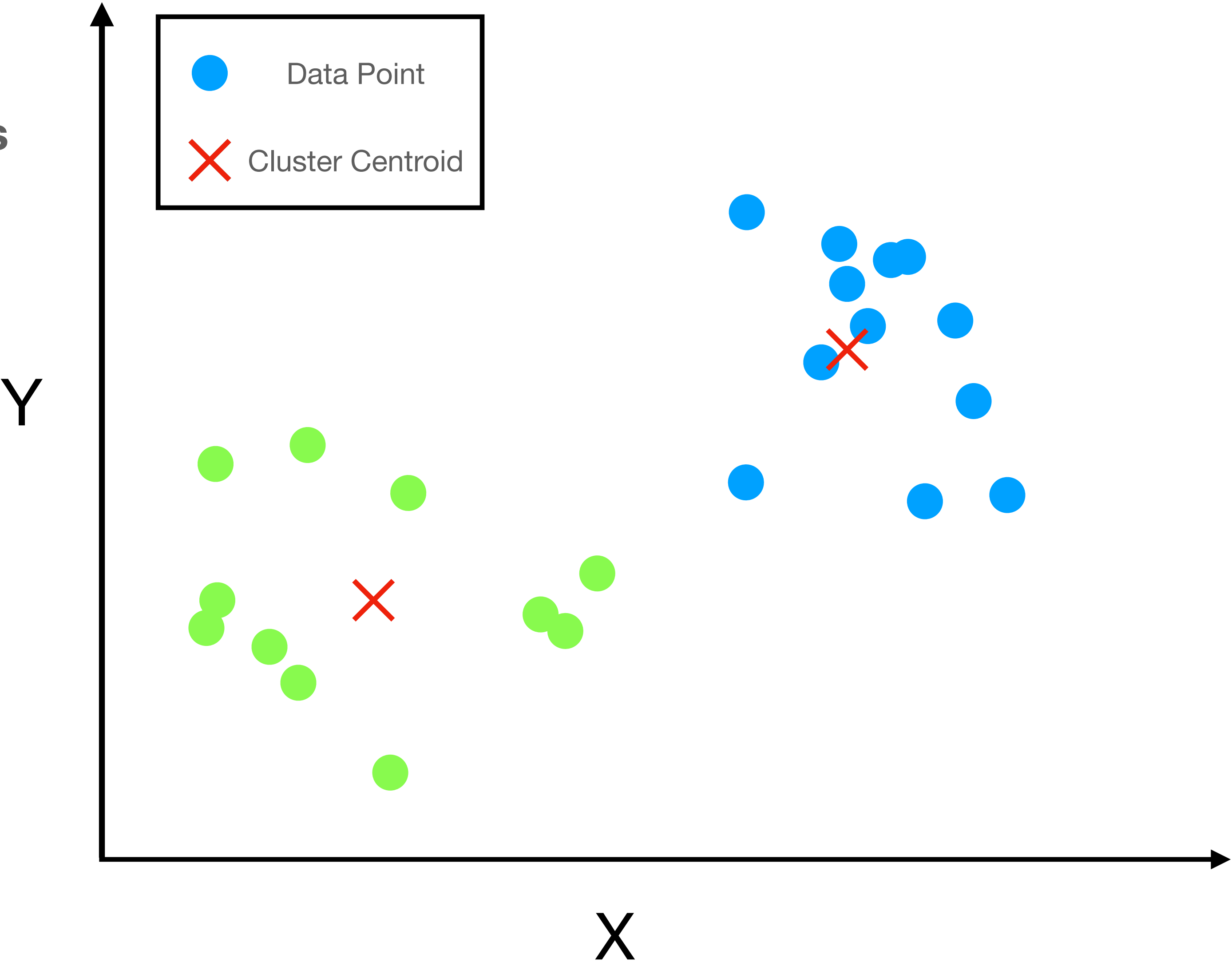
Step 1: Define centroids
Step 2: Cluster data



Clustering: K-means

Lloyd, 1982, IEEE

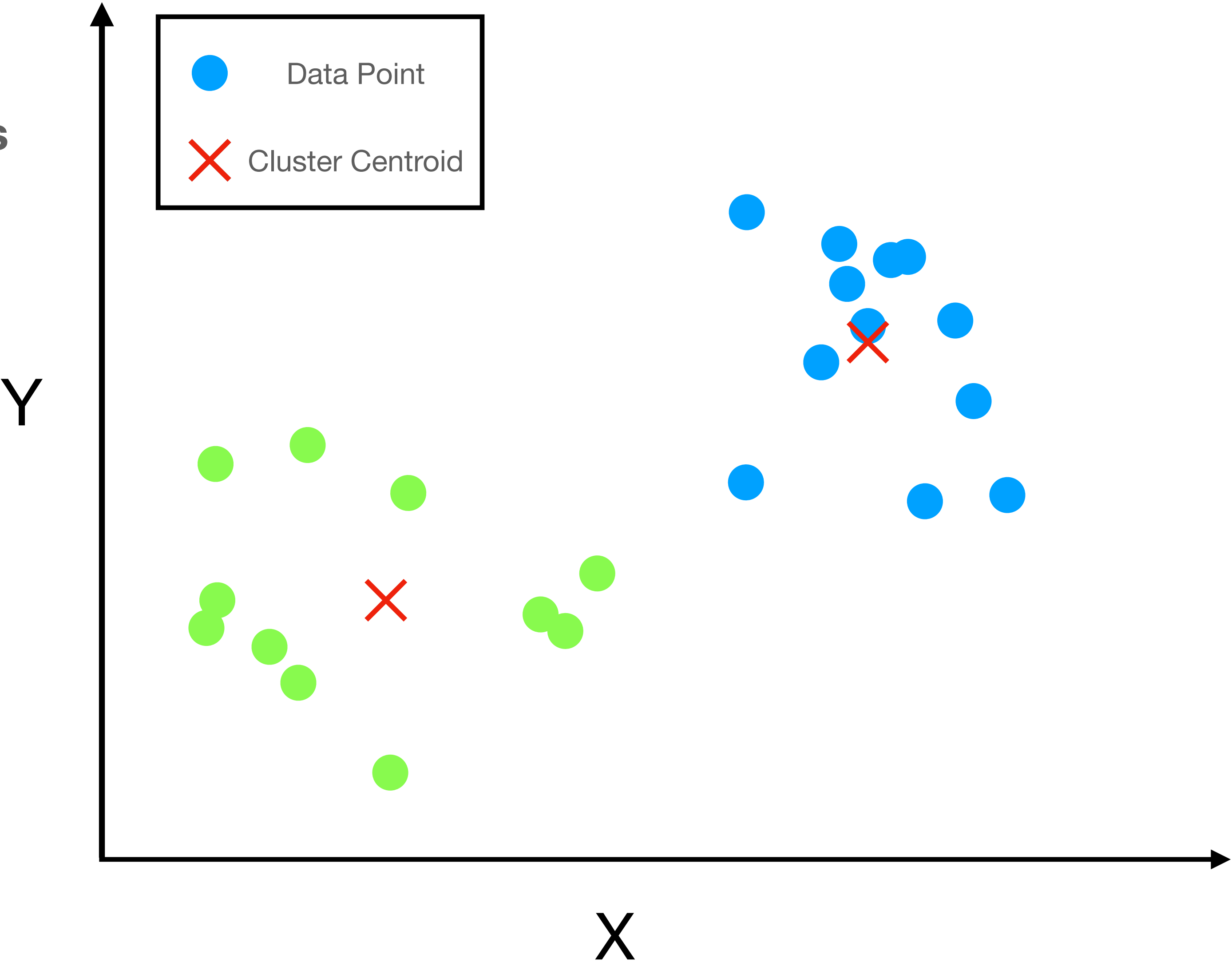
Step 1: Define centroids
Step 2: Cluster data



Clustering: K-means

Lloyd, 1982, IEEE

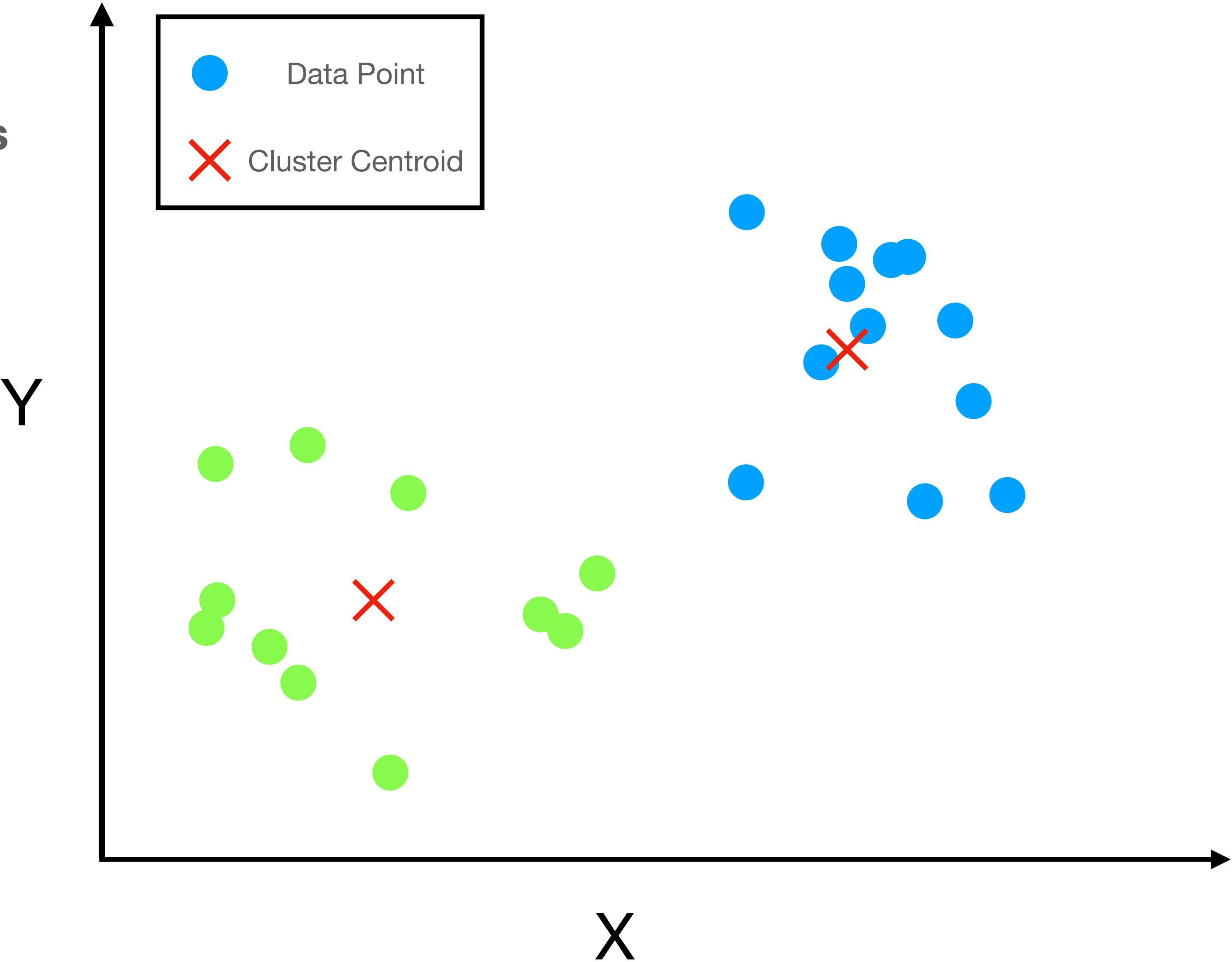
Step 1: Define centroids
Step 2: Cluster data



Clustering: K-means

Lloyd,1982, IEEE

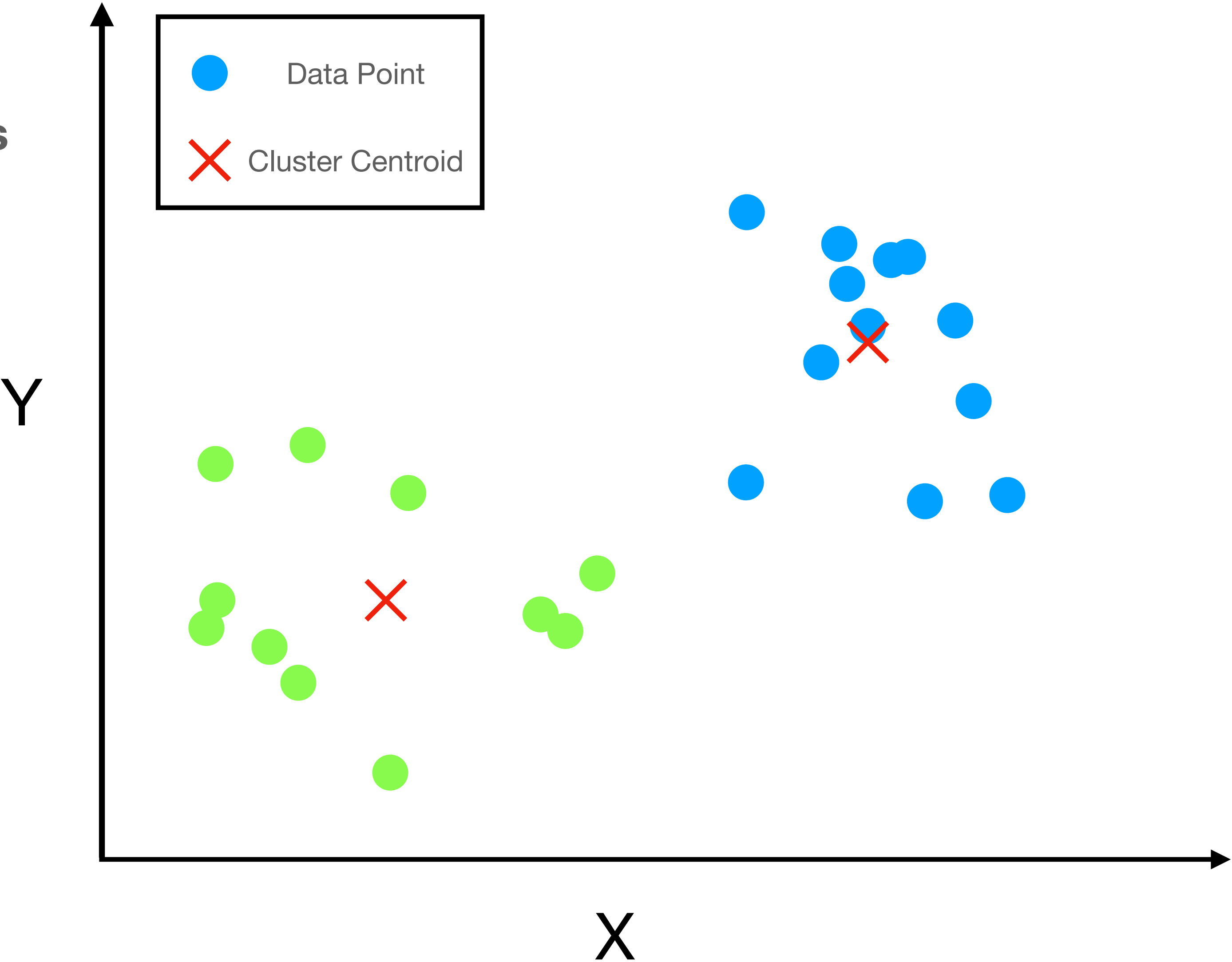
Step 1: Define centroids
Step 2: Cluster data



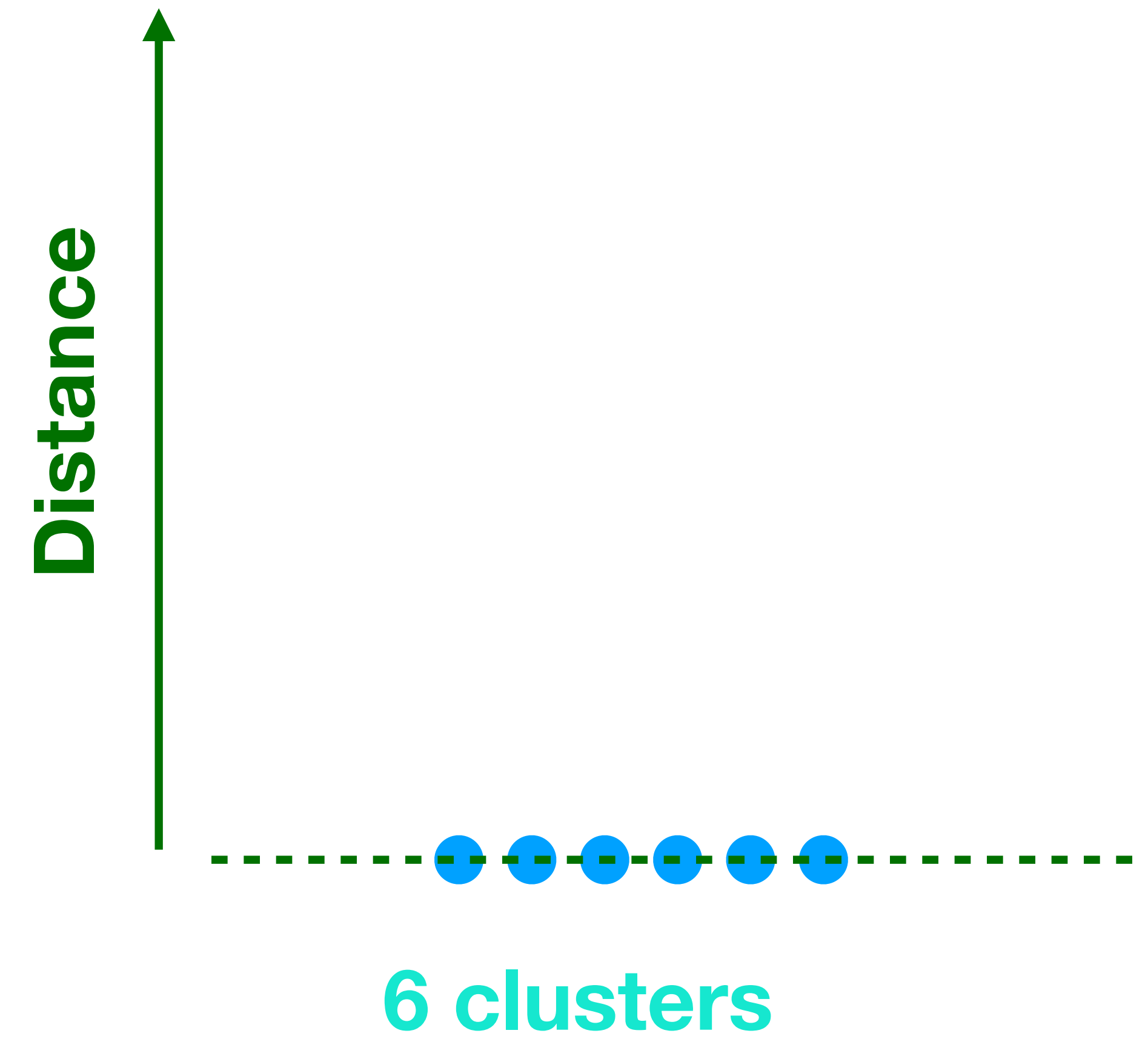
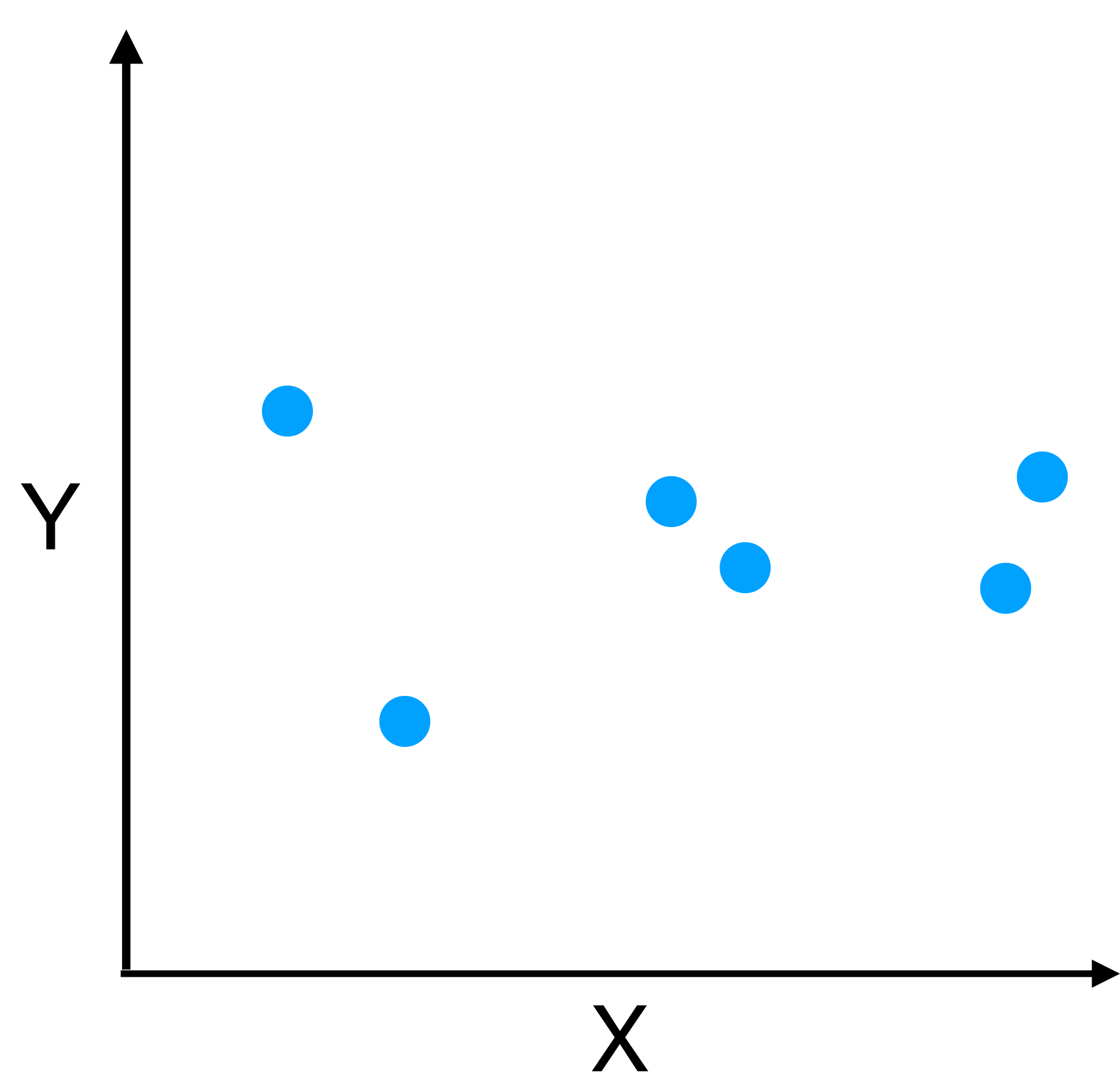
Clustering: K-means

Lloyd, 1982, IEEE

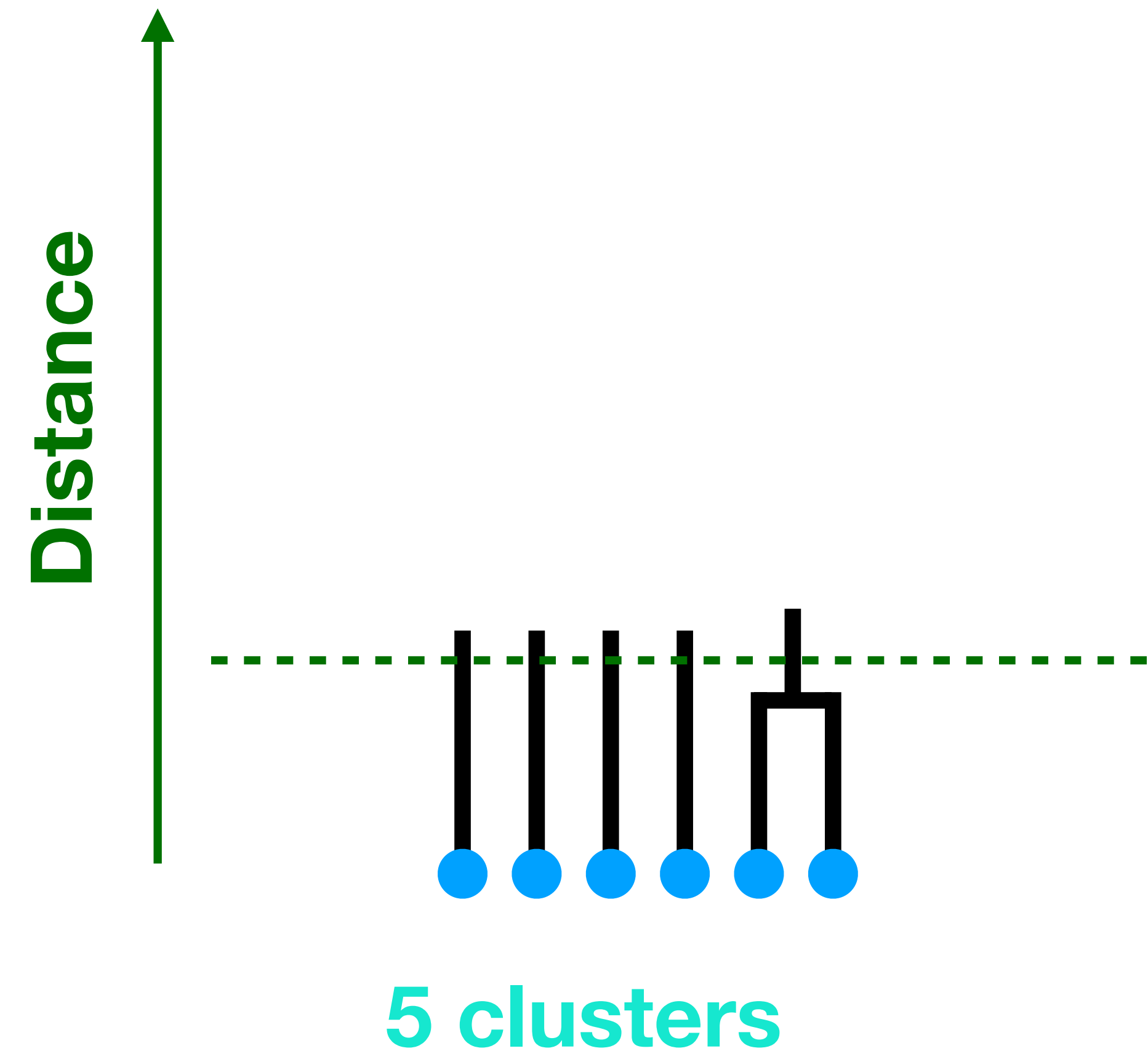
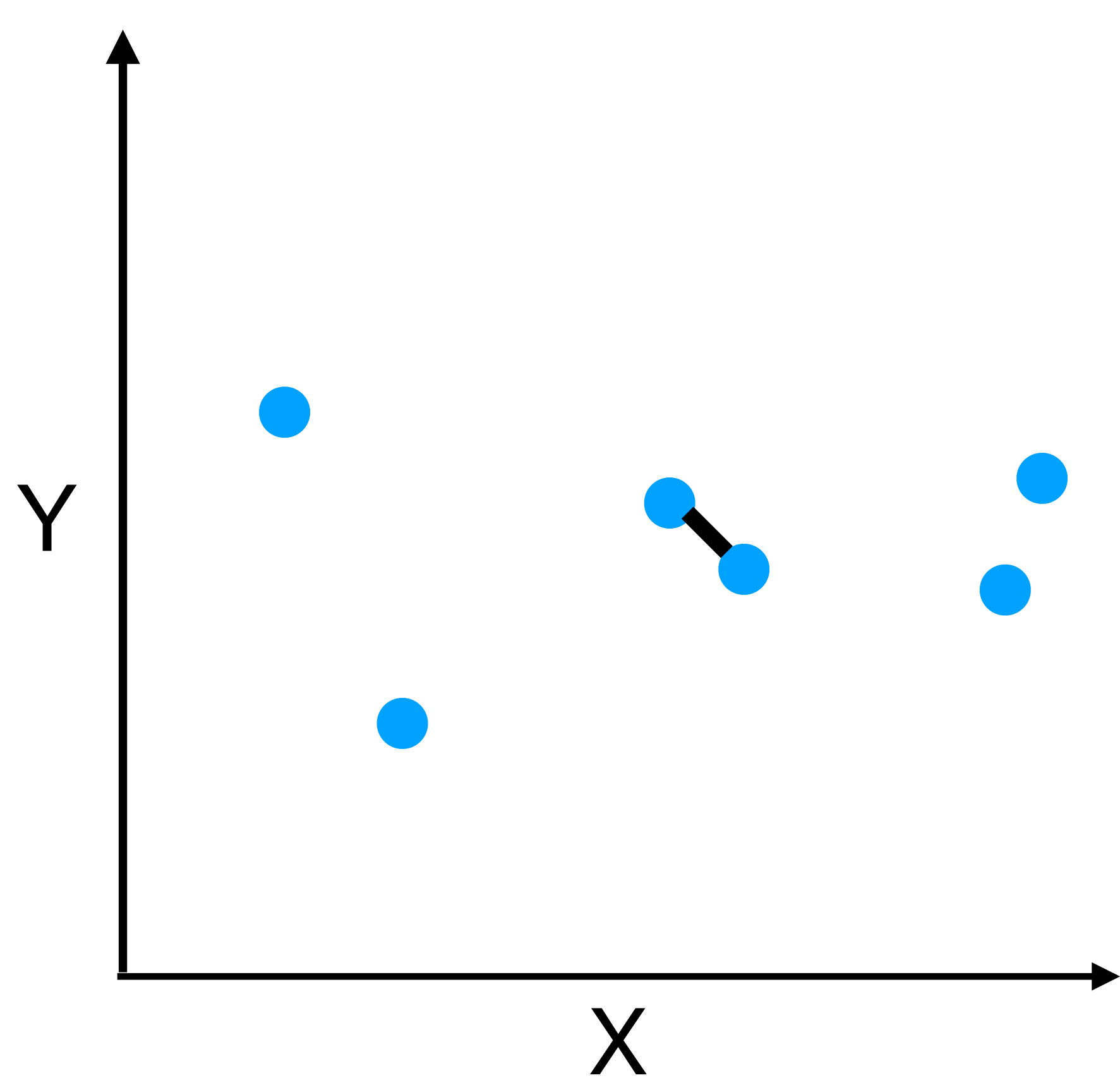
Step 1: Define centroids
Step 2: Cluster data



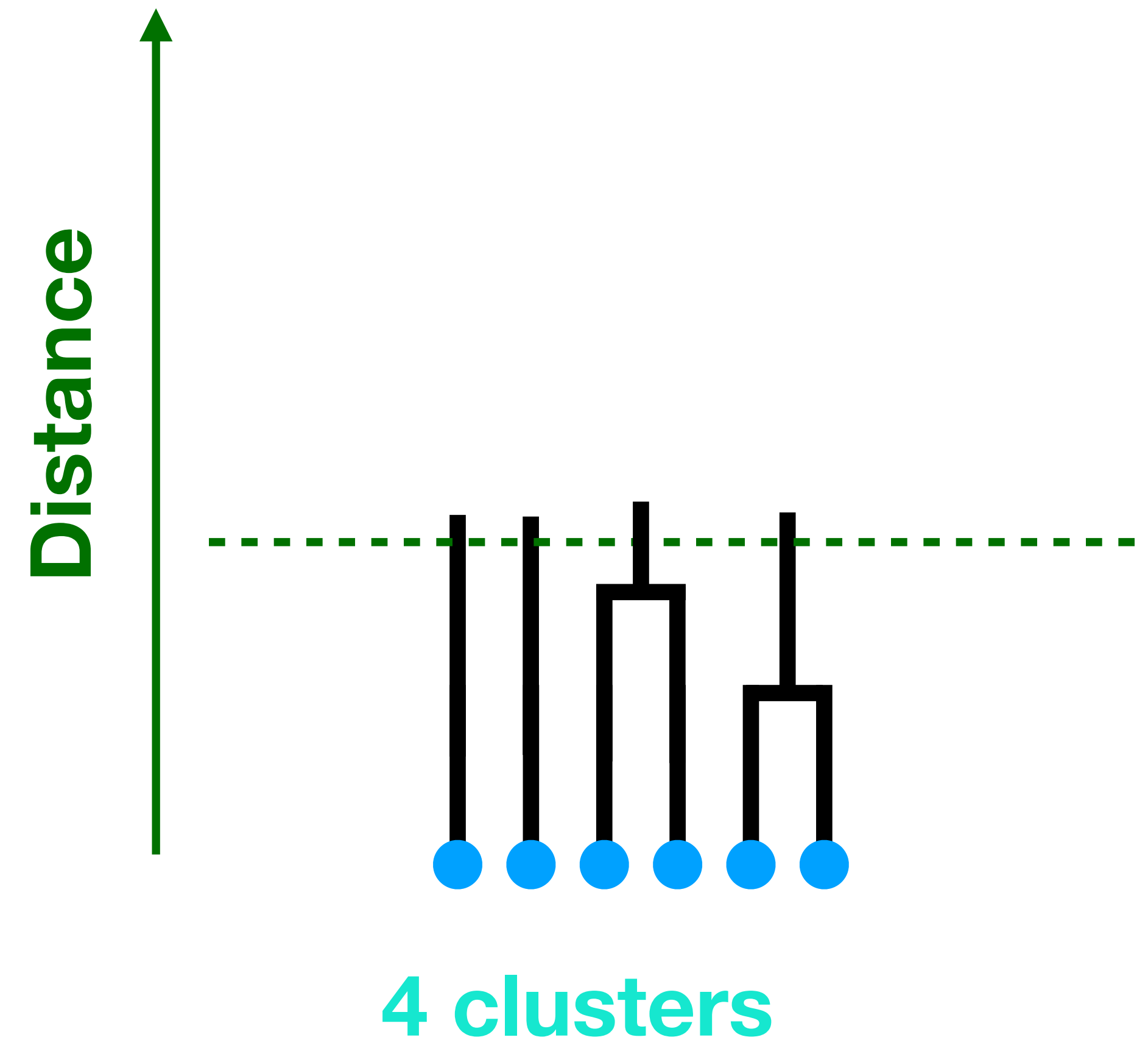
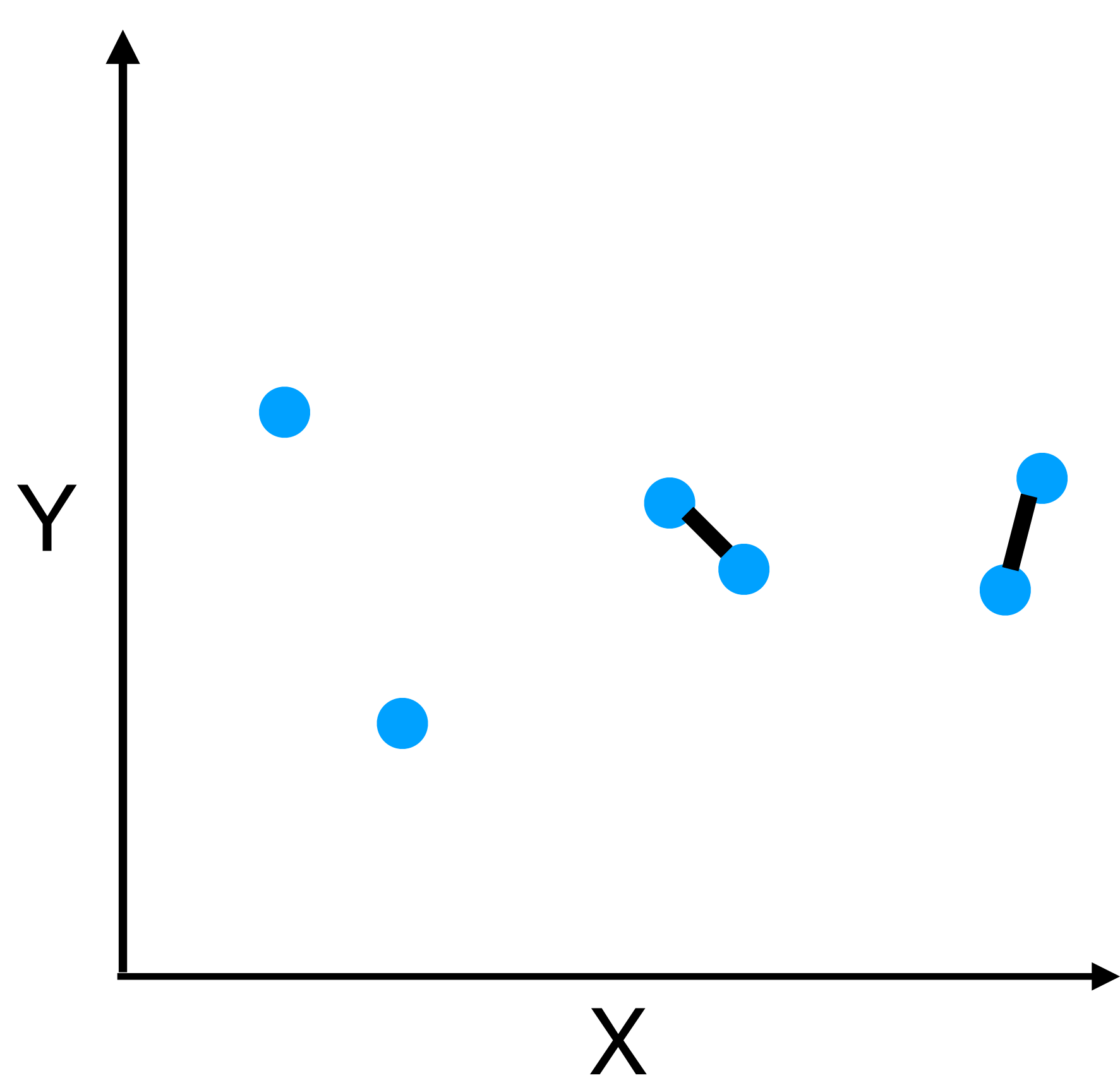
Hierarchical Clustering w/ Single Linkage



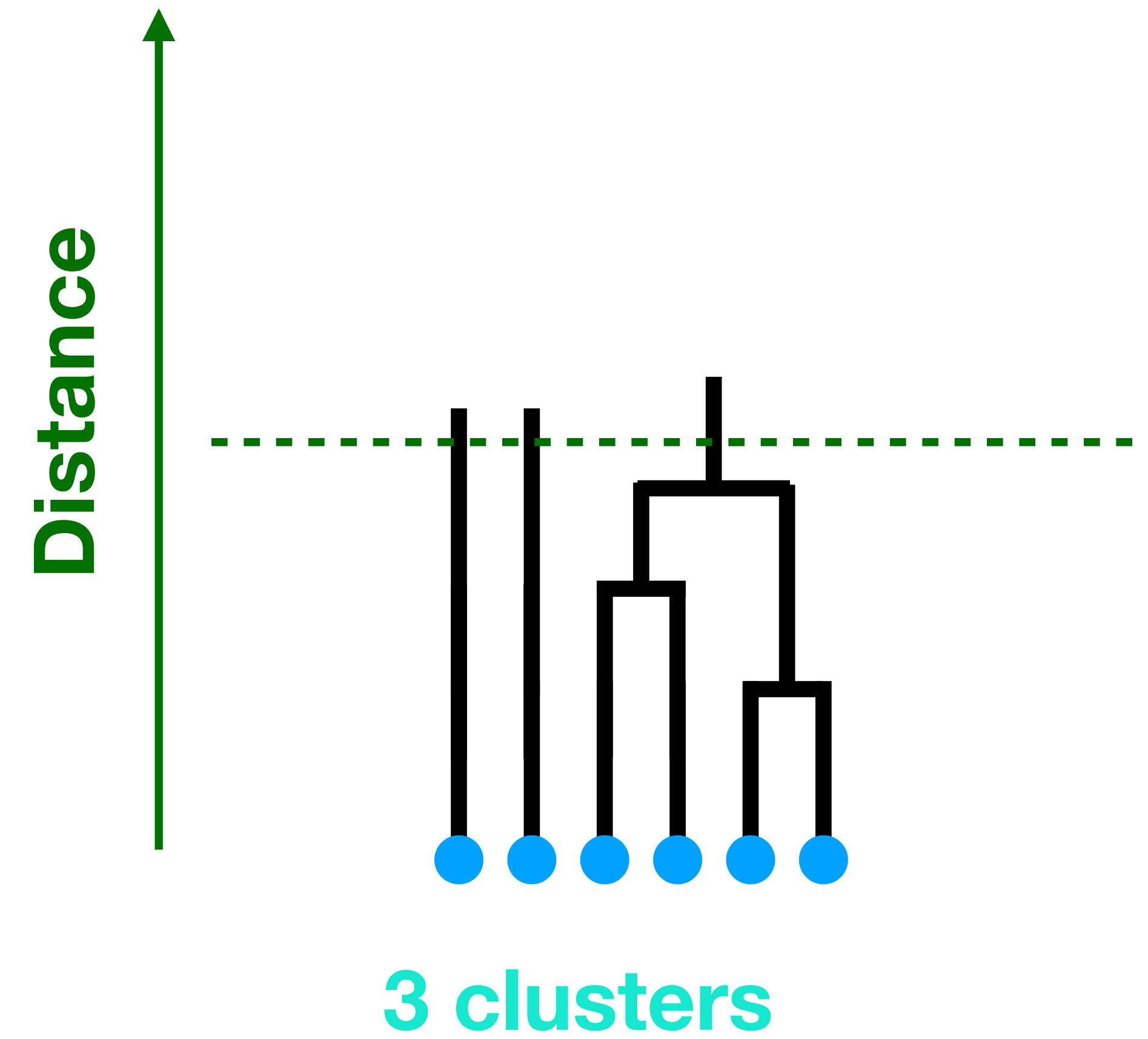
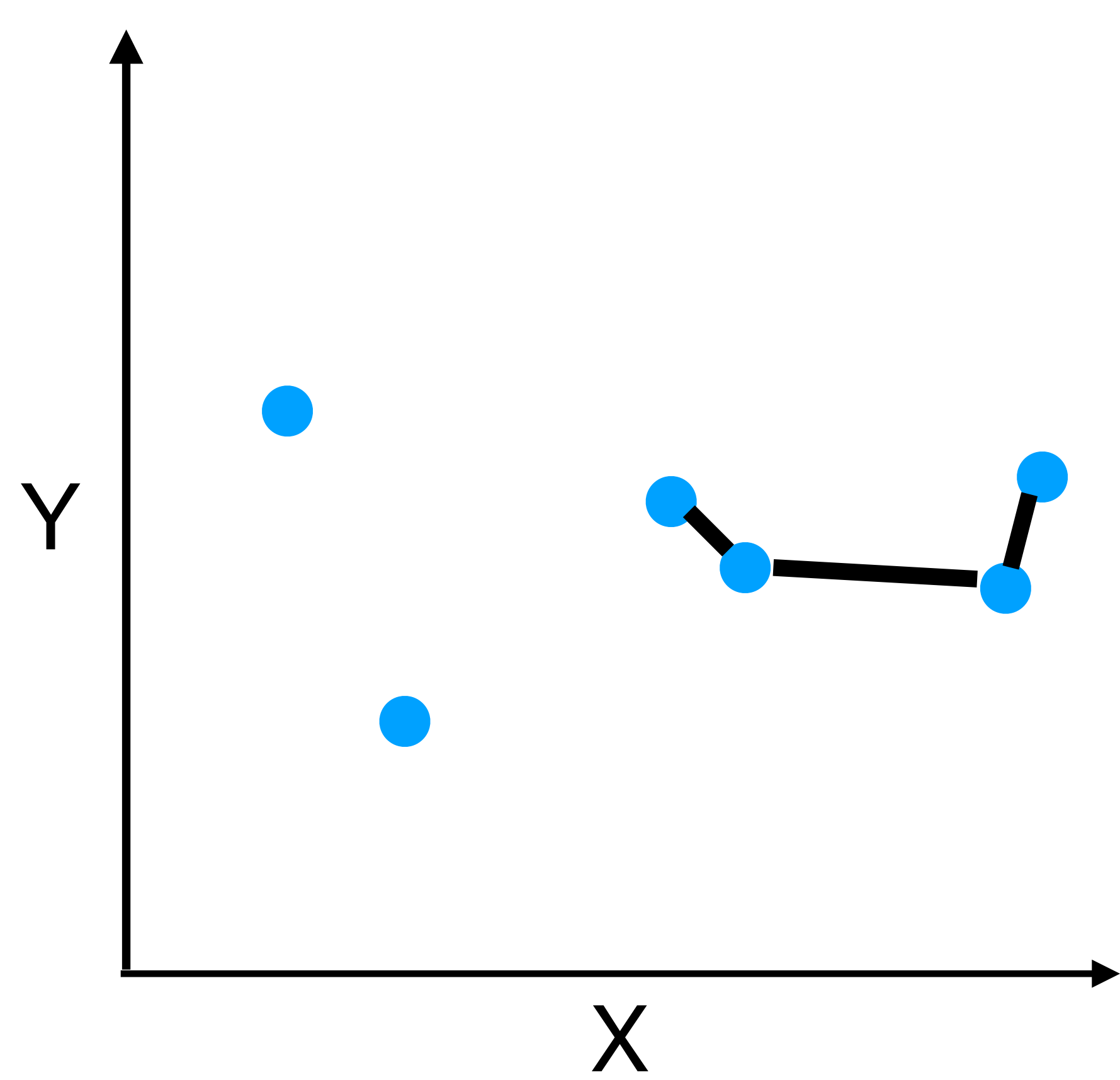
Hierarchical Clustering w/ Single Linkage



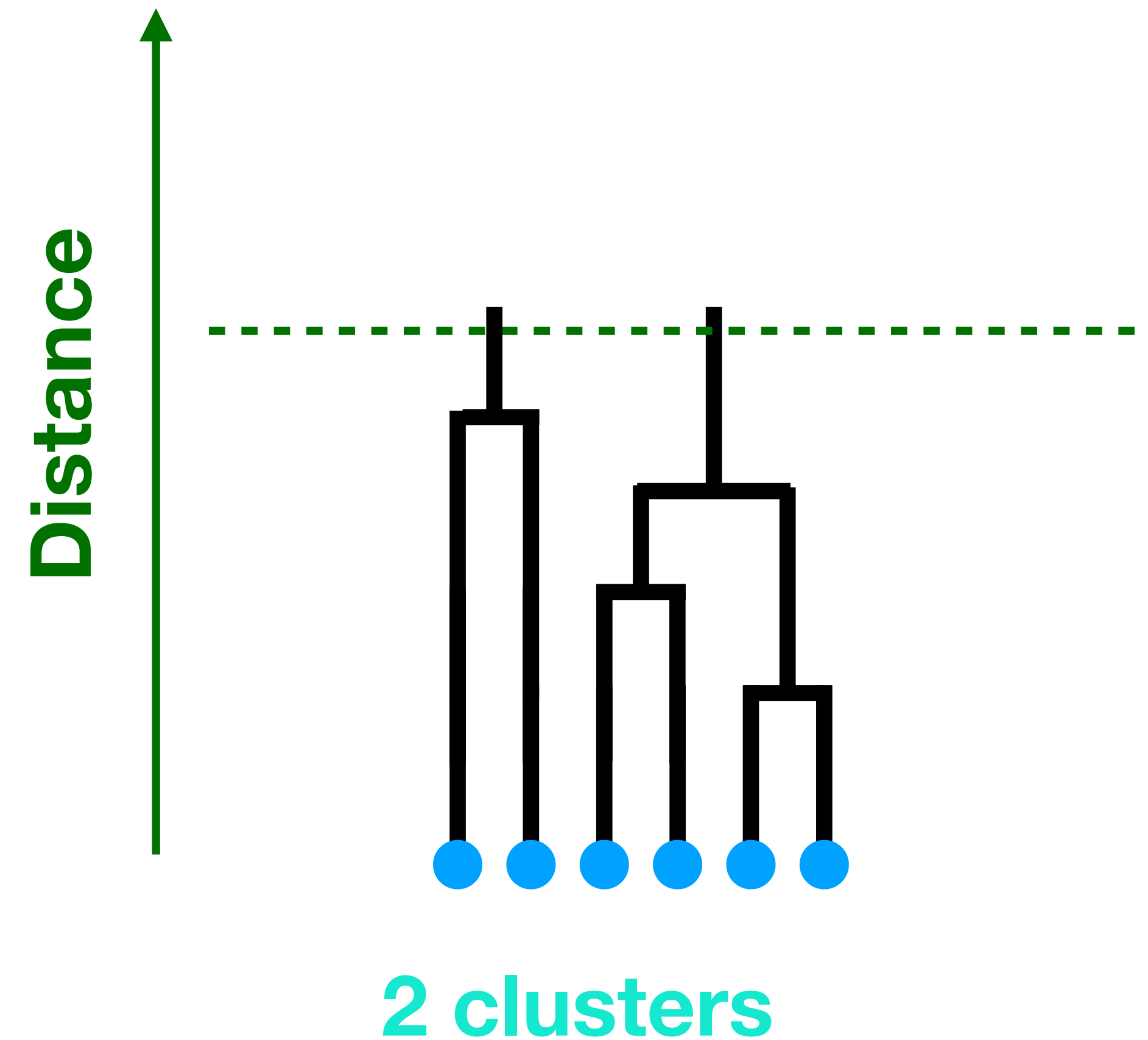
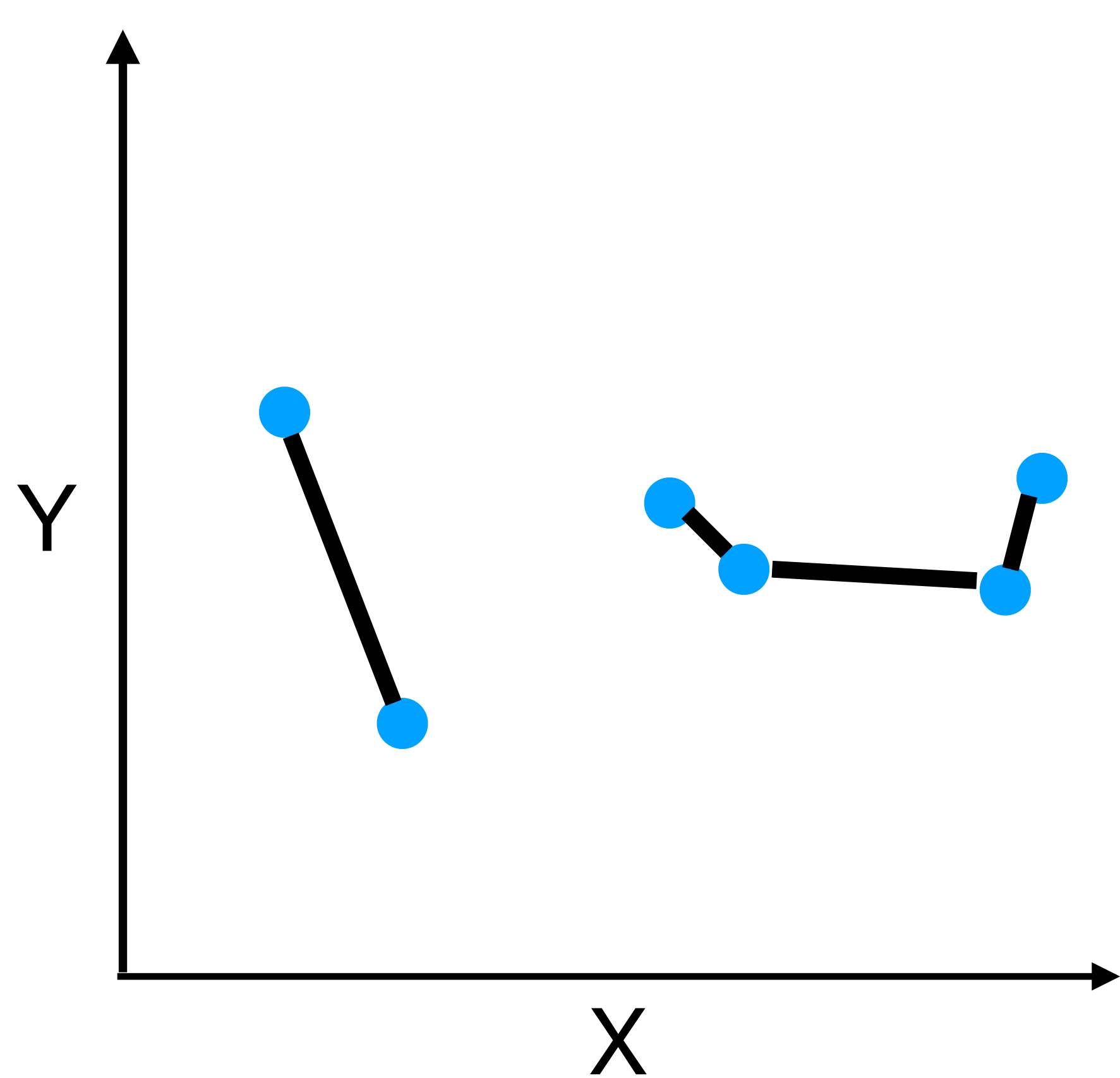
Hierarchical Clustering w/ Single Linkage



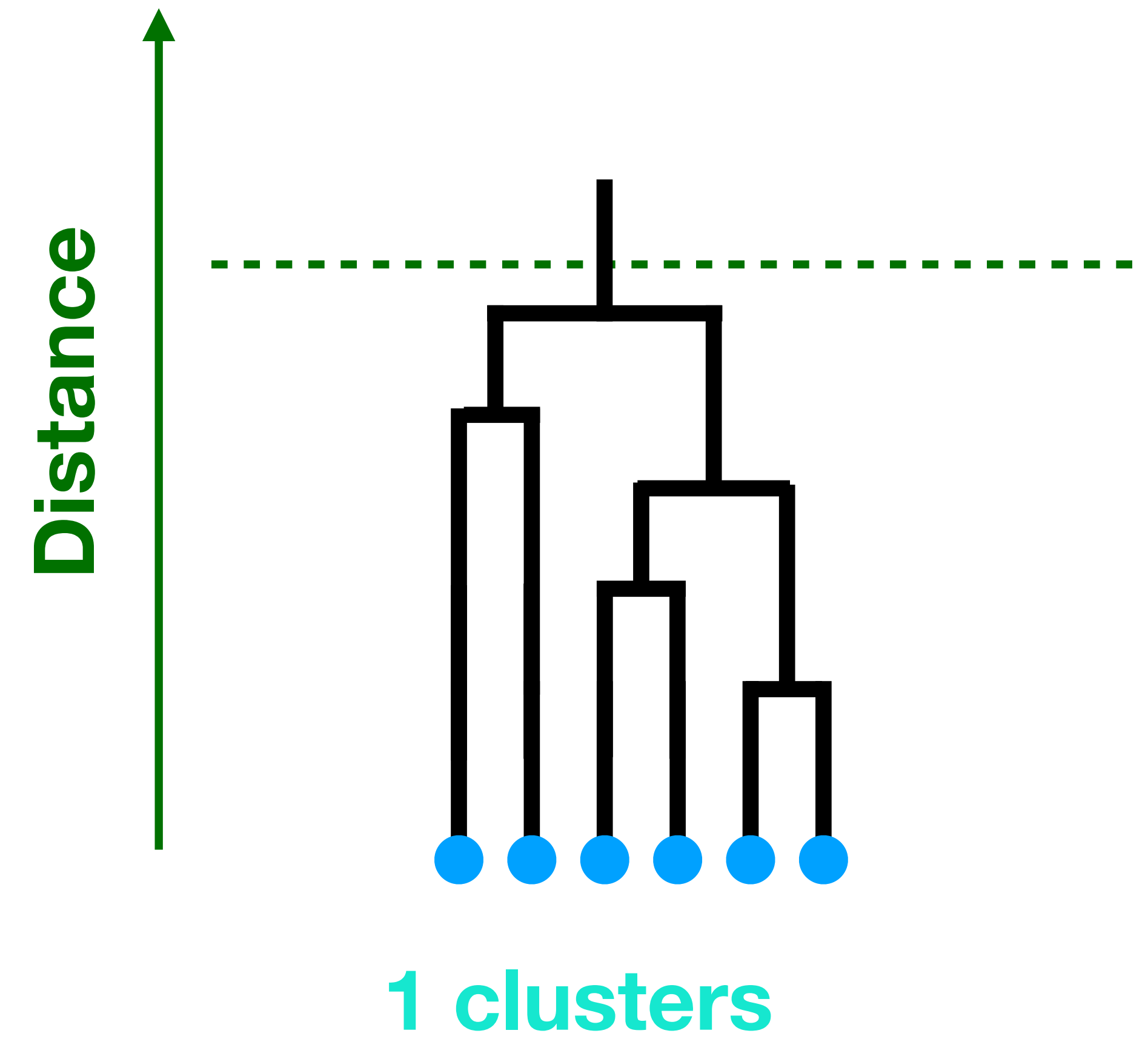
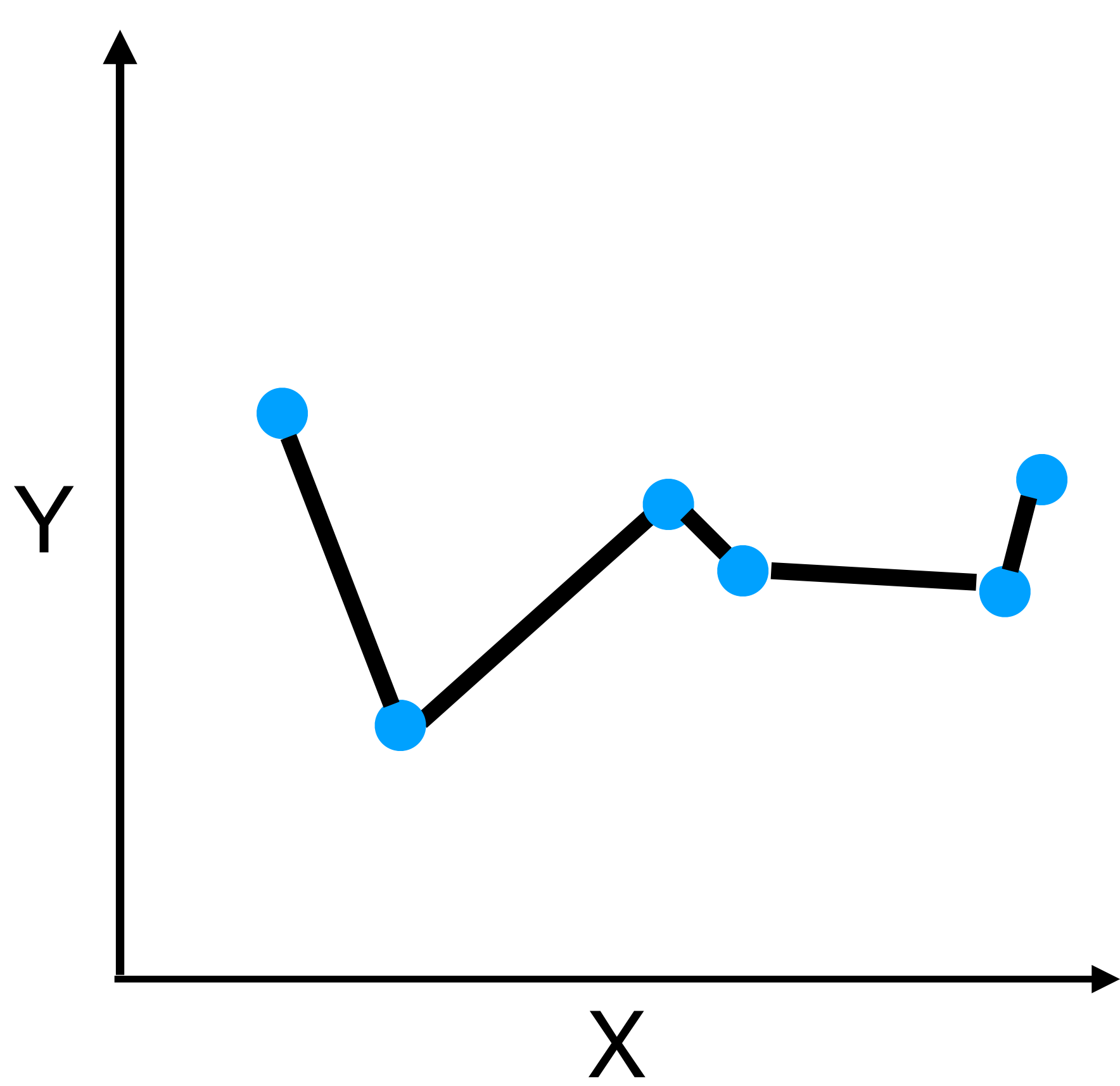
Hierarchical Clustering w/ Single Linkage



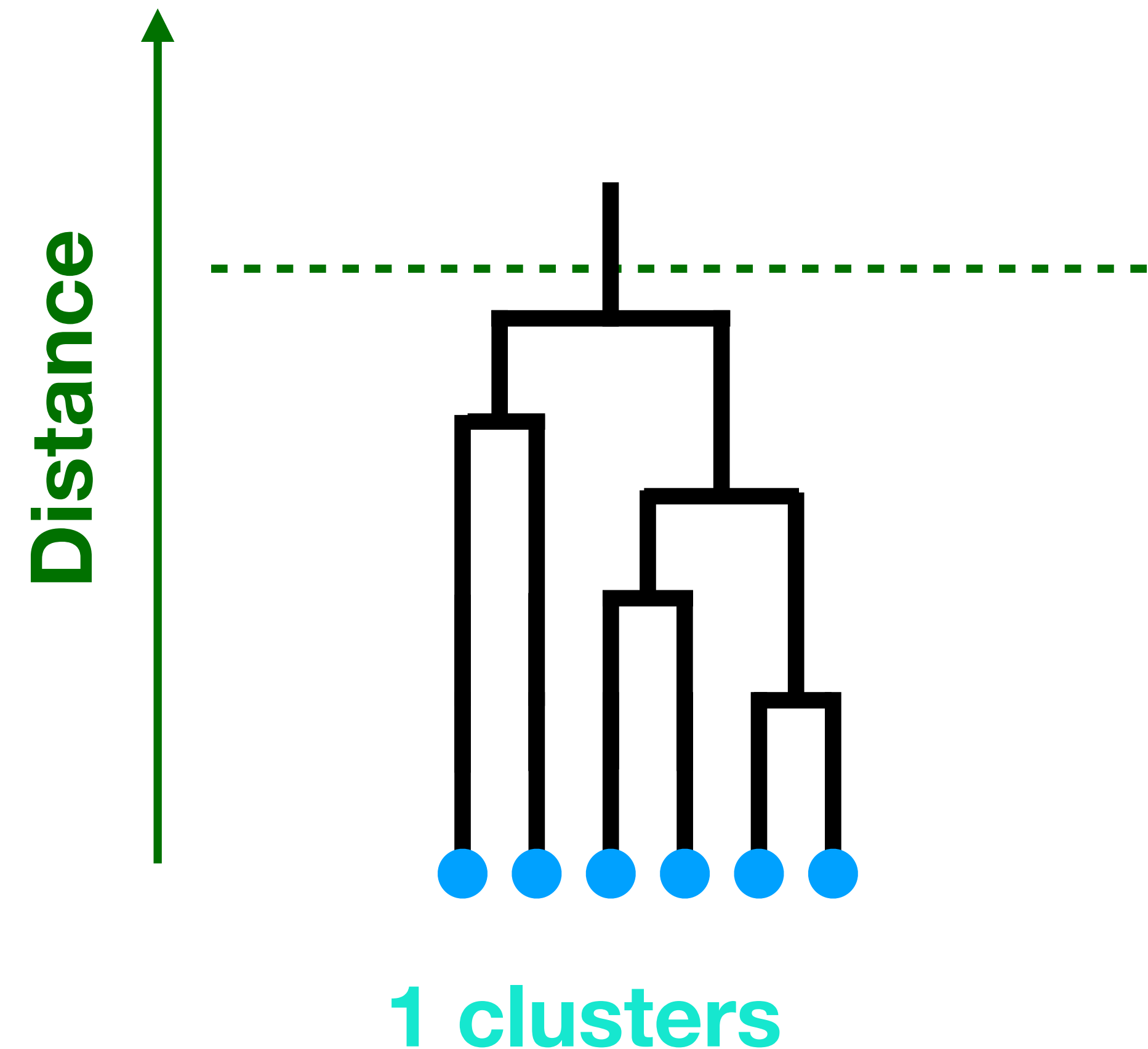
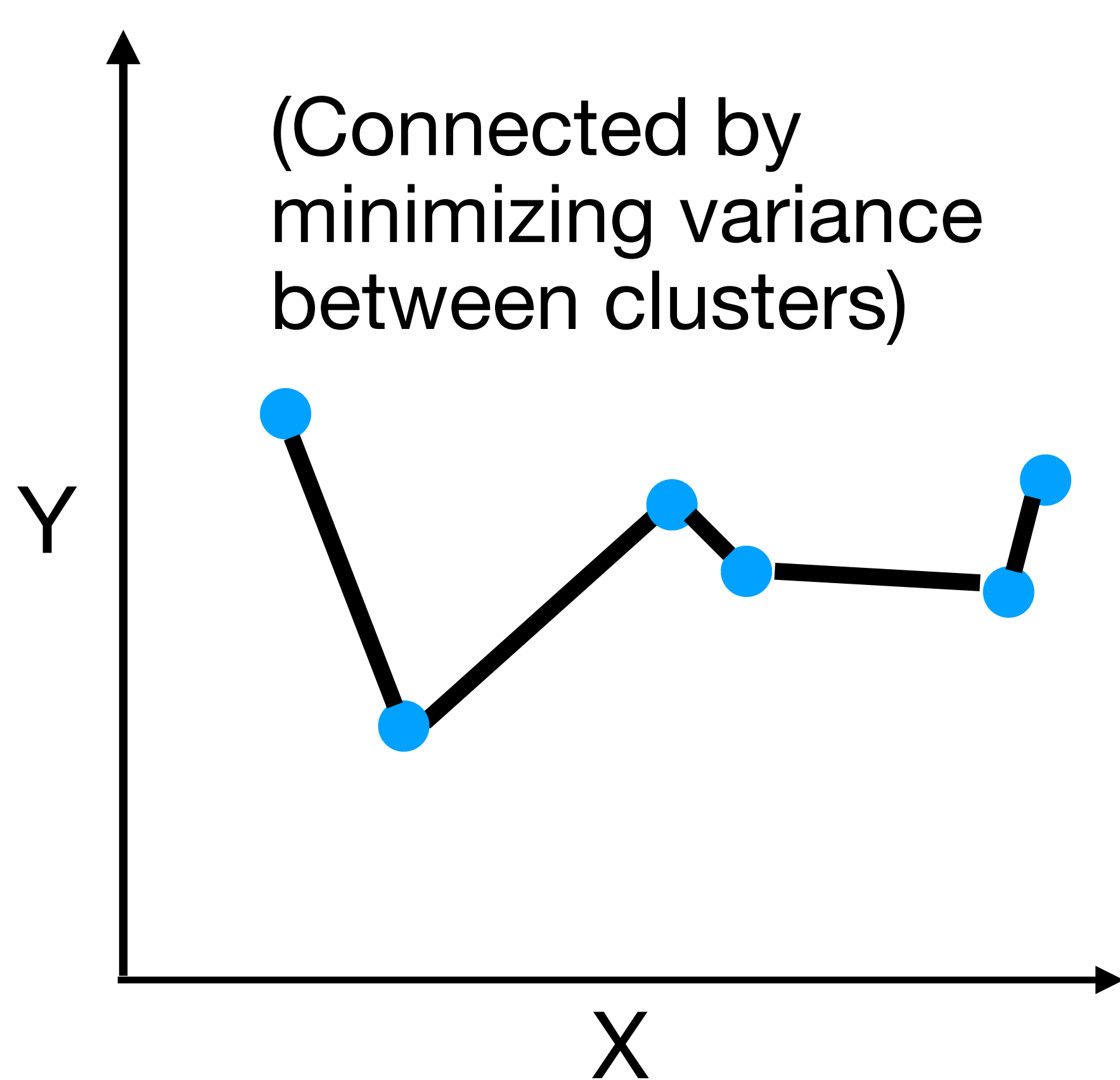
Hierarchical Clustering w/ Single Linkage



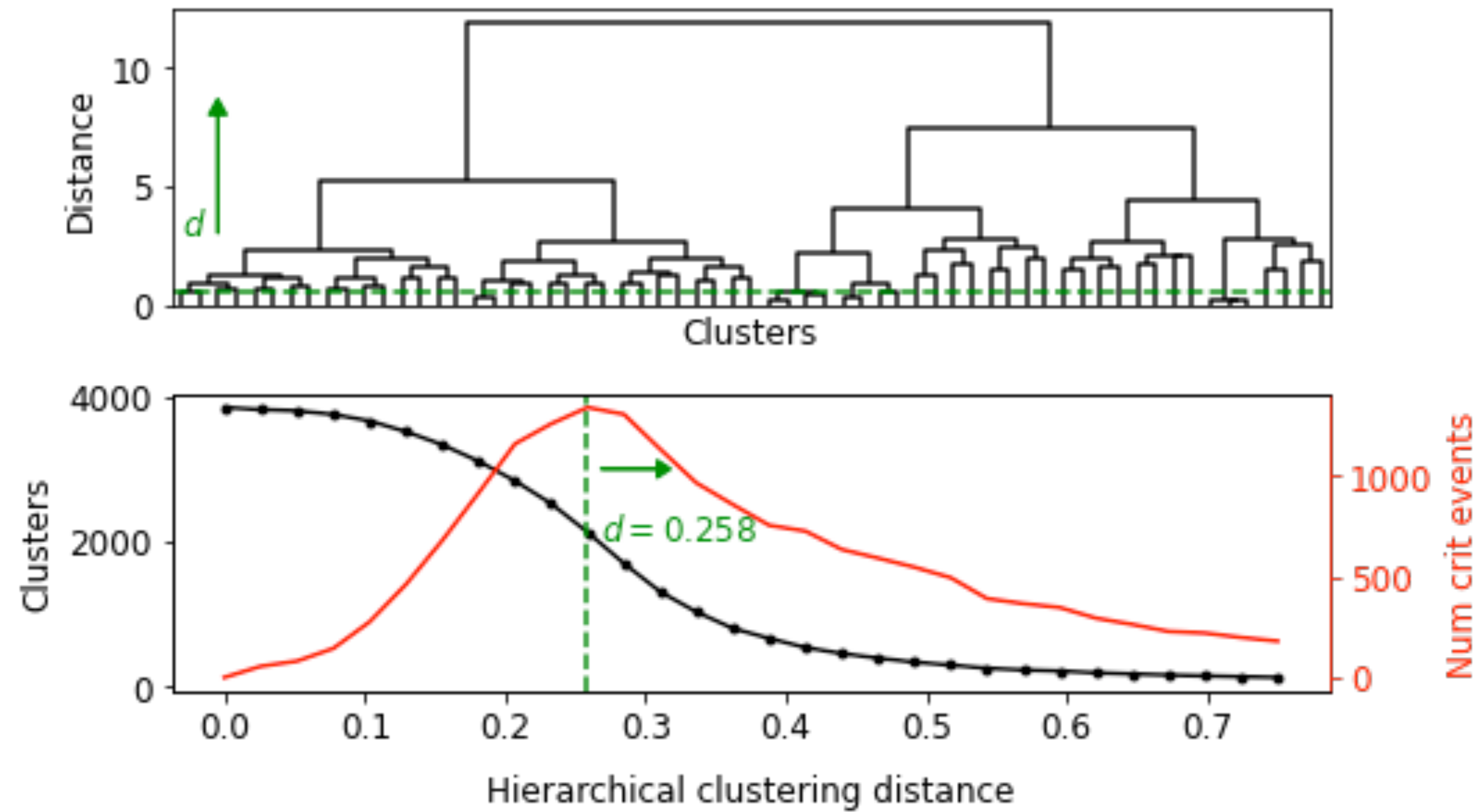
Hierarchical Clustering w/ Single Linkage



Hierarchical Clustering w/ Ward Linkage



How many clusters: Dendrogram



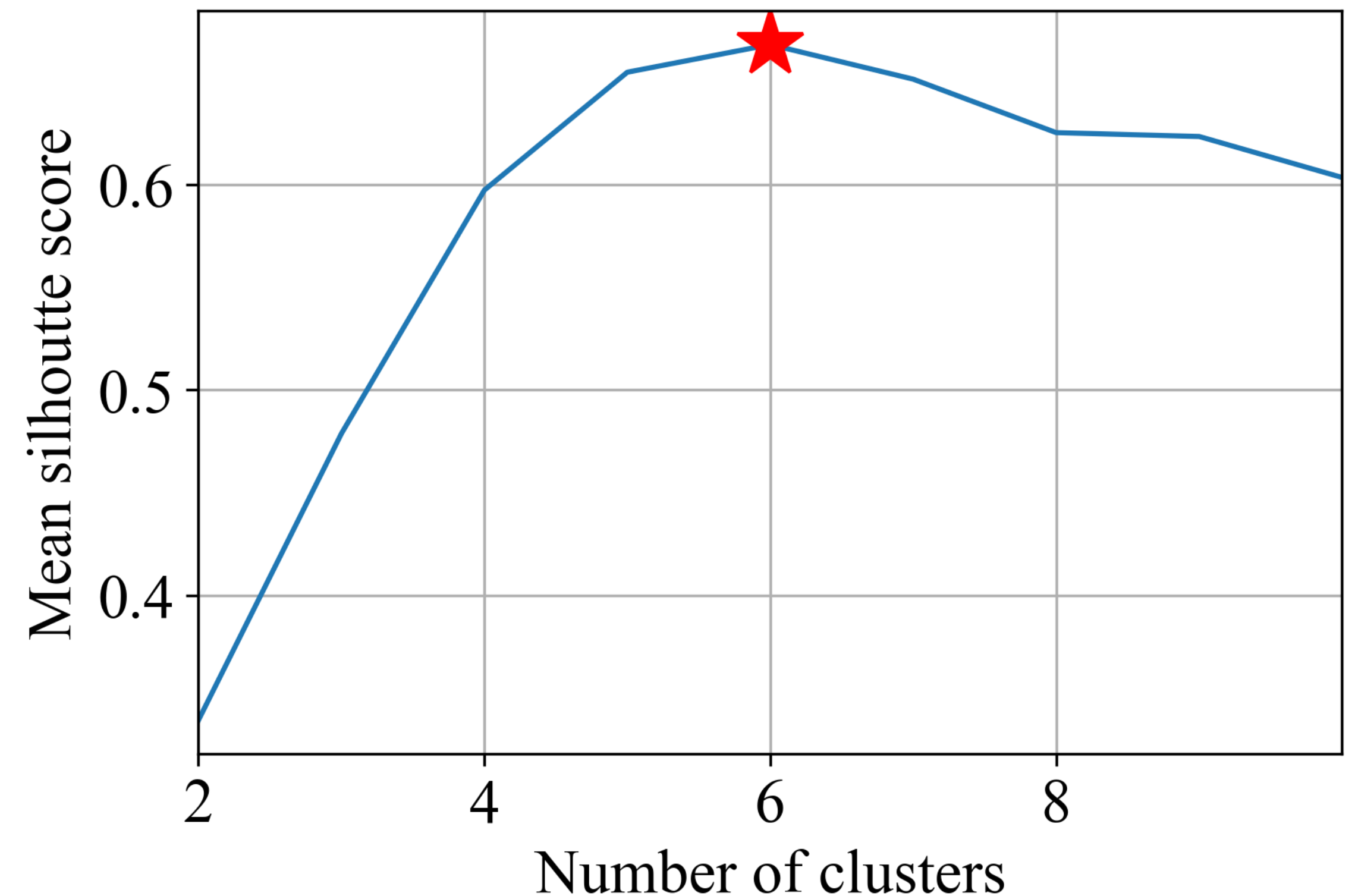
How many clusters: Silhouette Scores (SiS)

a : average Euclidean distance of x to every other data point in its own cluster

b : average Euclidean distance of x to every data point in the closest neighboring cluster

$$SiS(x) = \frac{b - a}{\max(a, b)}$$

Rousseeuw 1987



Summary

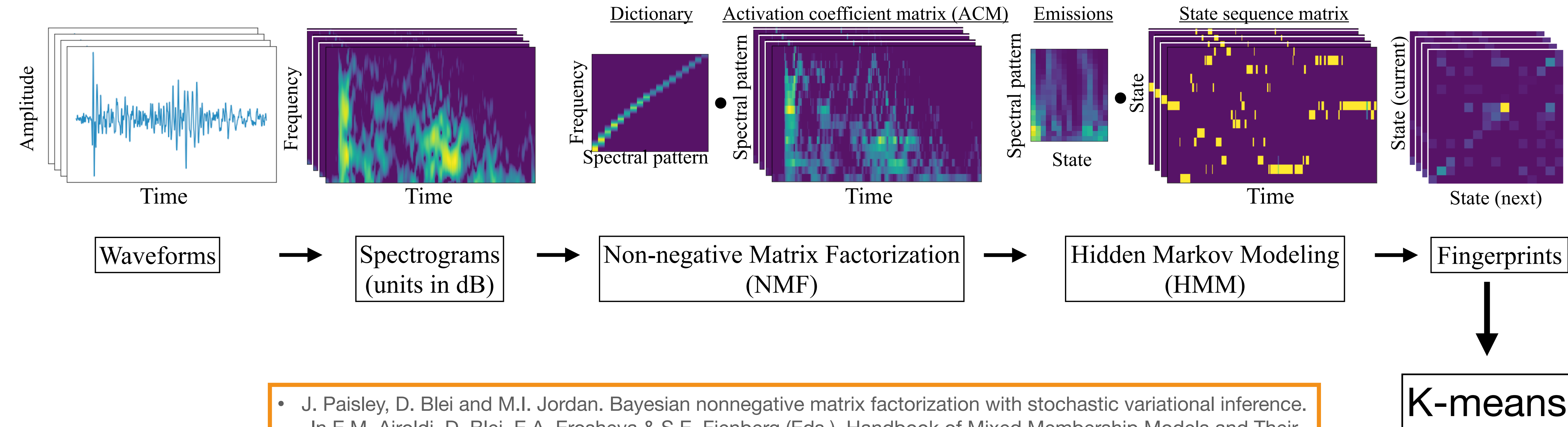
- Unsupervised ML infers patterns in large data sets without the need for prior training labels
- Feature extraction and dimensionality reduction can aid clustering and interpretation
 - *Examples: Nonnegative matrix factorization (NMF), K-means clustering, hierarchical clustering*

Outline

- Introduction to unsupervised machine learning (UML)
 - feature extraction & dimensionality reduction
 - clustering
- **SpecUFEx tutorial: Amatrice 2016**

Spectral Unsupervised Feature Extraction (SpecUFEx)

Holtzman et al., 2018; Sci Adv

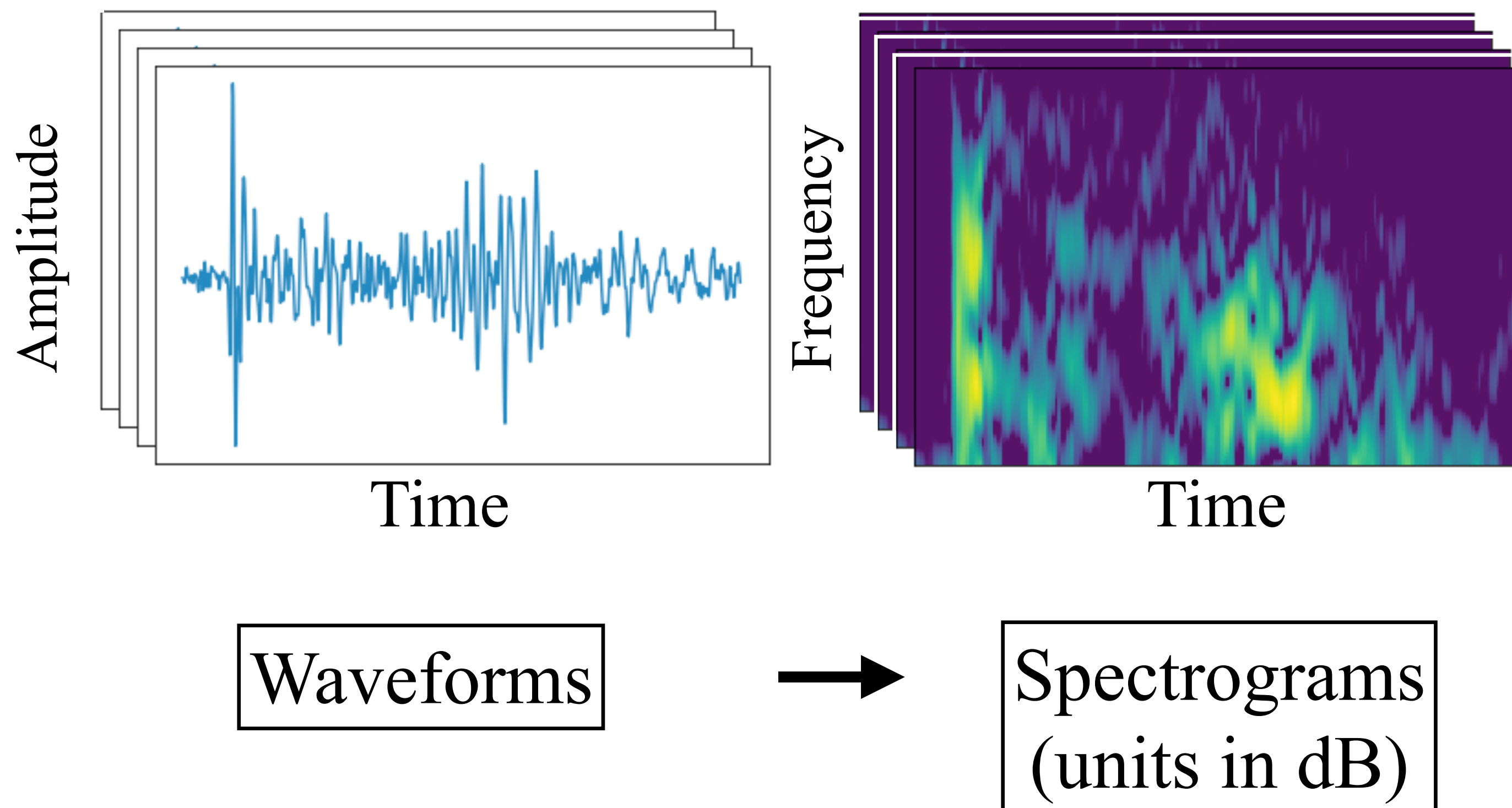


- J. Paisley, D. Blei and M.I. Jordan. Bayesian nonnegative matrix factorization with stochastic variational inference. In E.M. Airoldi, D. Blei, E.A. Erosheva & S.E. Fienberg (Eds.), Handbook of Mixed Membership Models and Their Applications. Chapman and Hall/CRC Handbooks of Modern Statistical Methods, 2014
- M. Hoffman, D. Blei, C. Wang and J. Paisley. Stochastic variational inference, Journal of Machine Learning Research, vol. 14, pp. 1303-1347, 2013.

Spectral Unsupervised Feature Extraction (SpecUFEx)

Holtzman et al., 2018; Sci Adv

Generate spectrogram

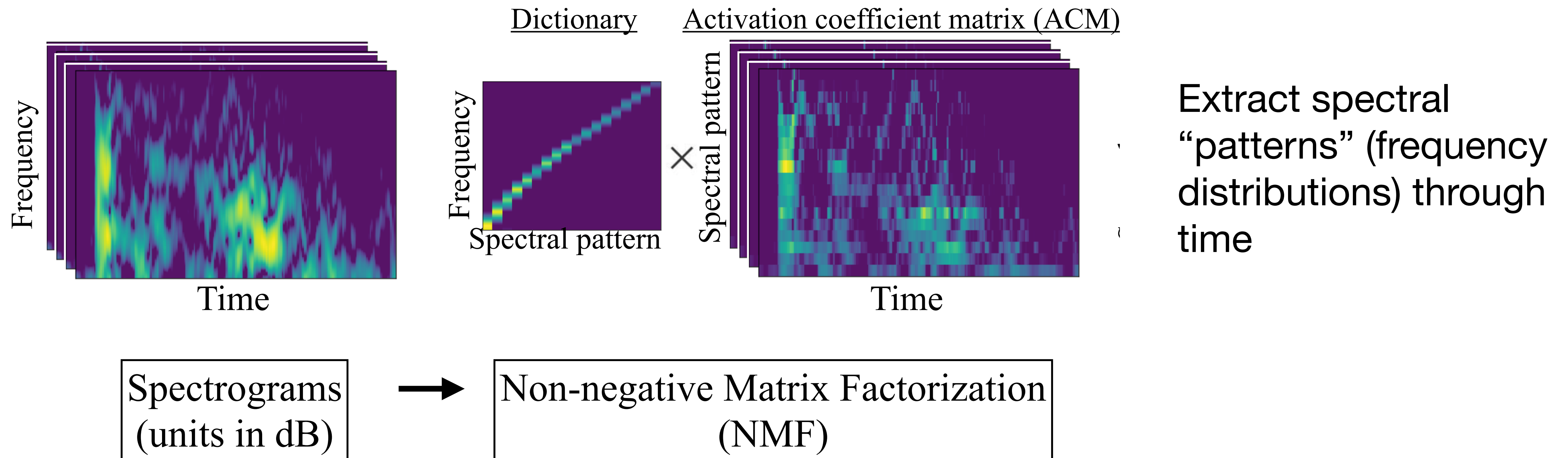


Ground amplitude through time ->
Frequency content through time

Spectral Unsupervised Feature Extraction (SpecUFEx)

Holtzman et al., 2018; Sci Adv

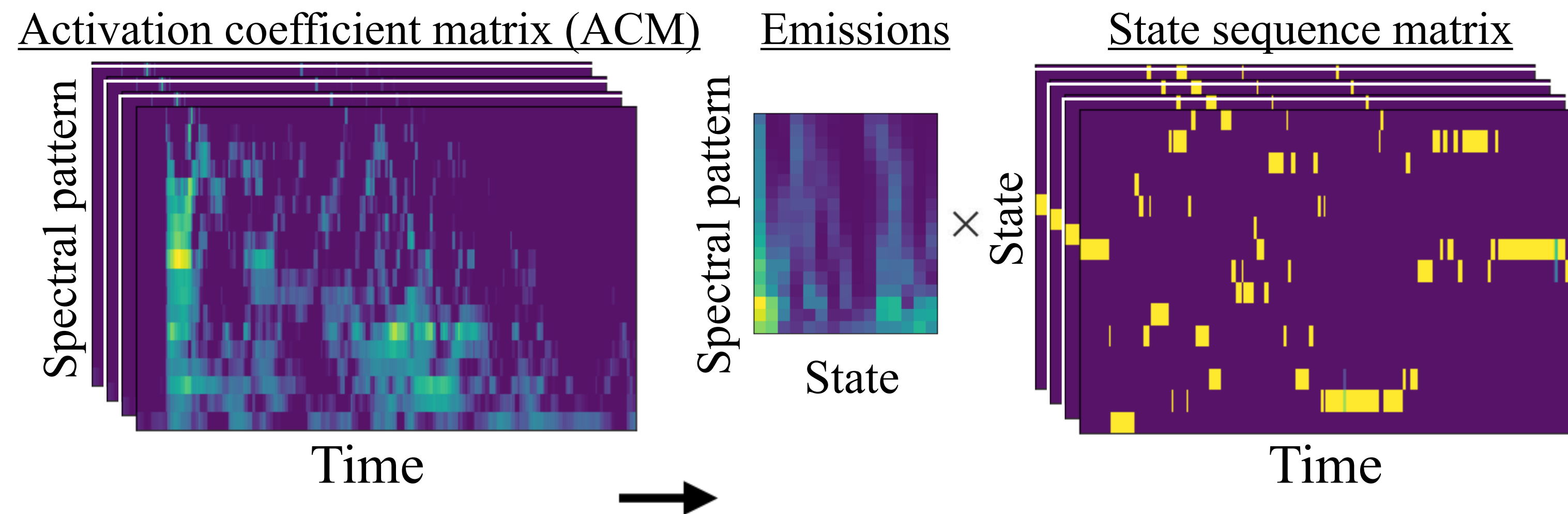
Non-negative Matrix Factorization



Spectral Unsupervised Feature Extraction (SpecUFEx)

Holtzman et al., 2018; Sci Adv

Hidden Markov Model

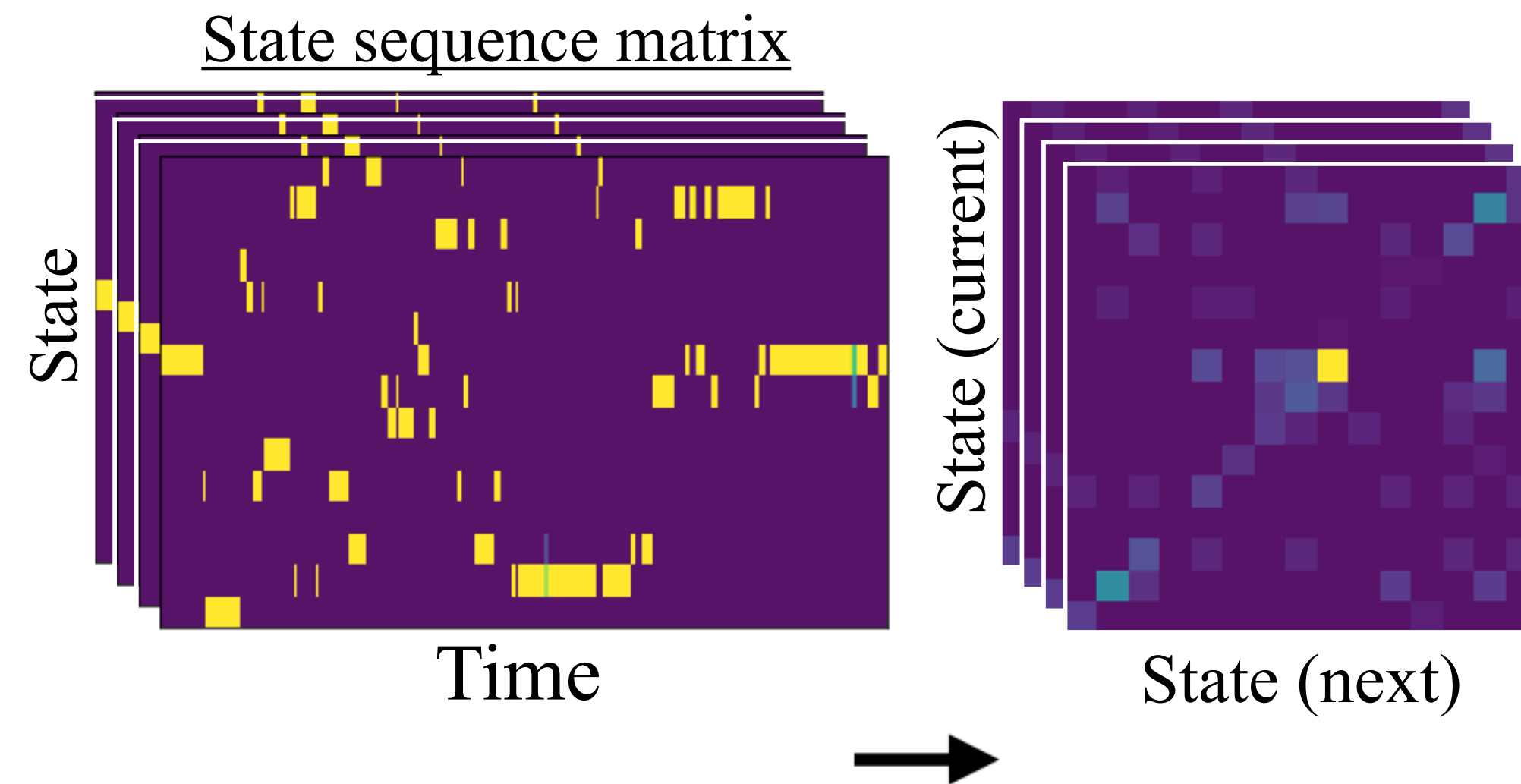


Hidden states control
concurrent spectral
patterns

Spectral Unsupervised Feature Extraction (SpecUFEx)

Holtzman et al., 2018; Sci Adv

Fingerprints

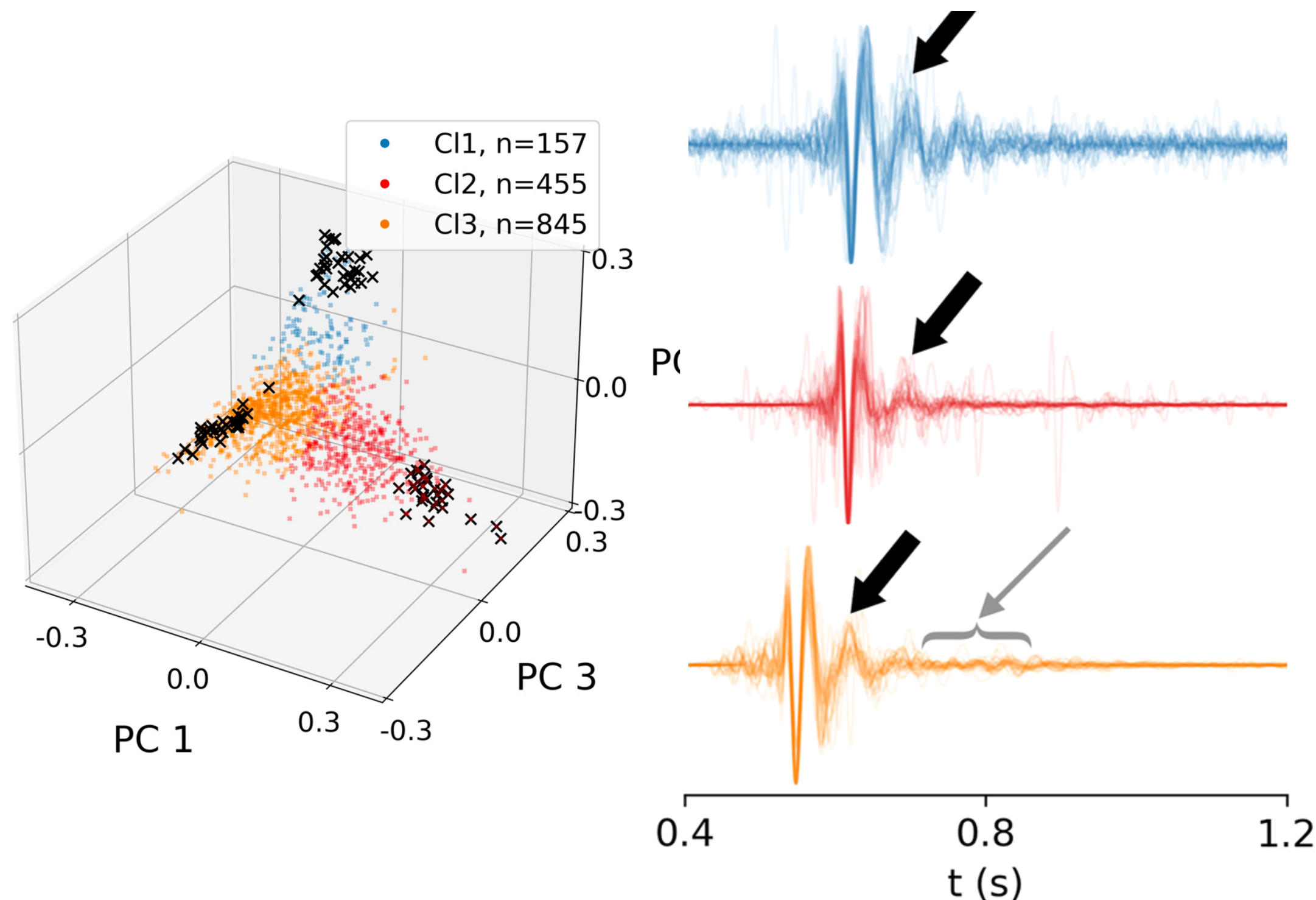


Count state transitions

Examples of SpecUFEx

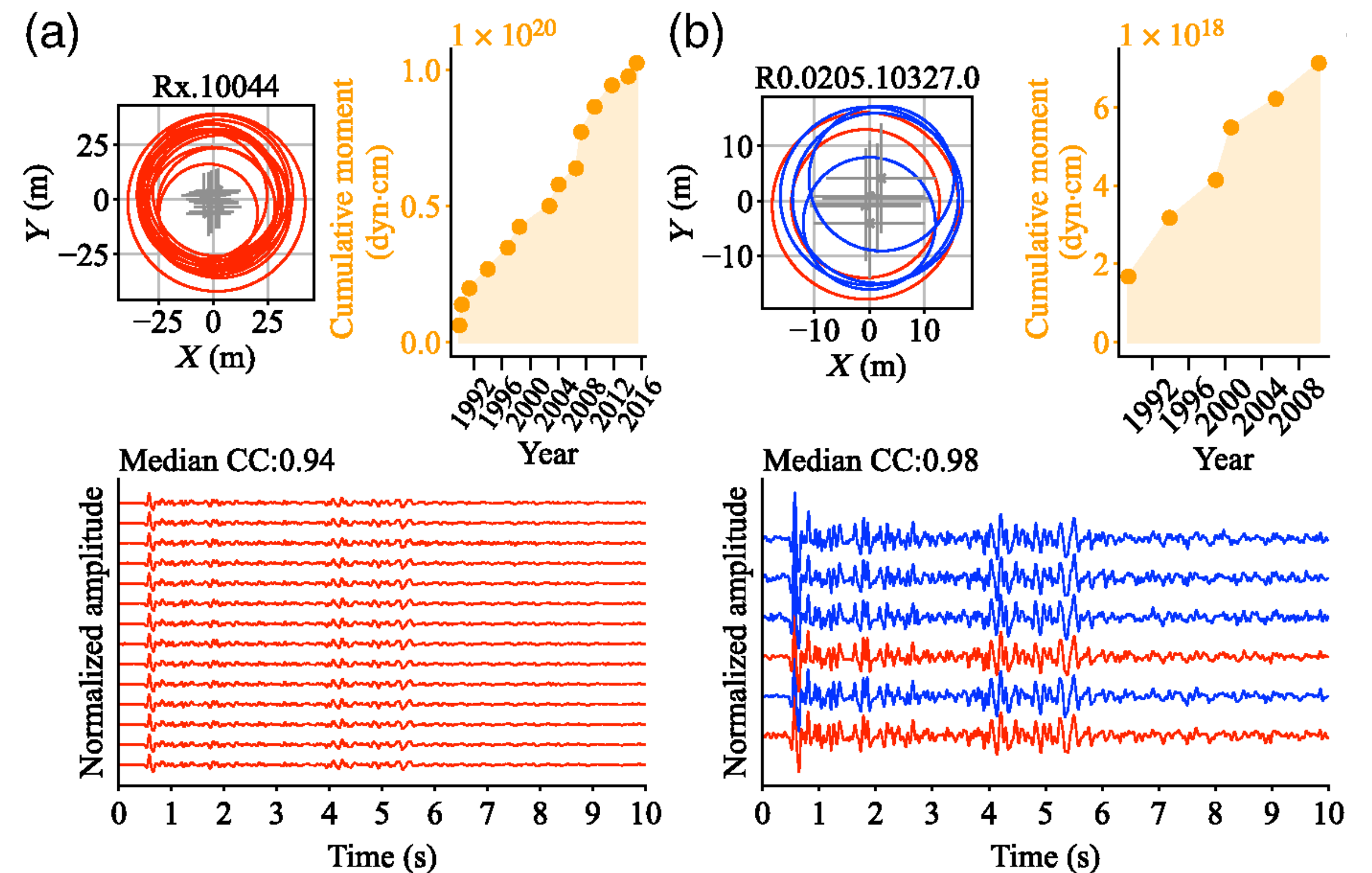
An unsupervised machine-learning approach to understanding seismicity at an alpine glacier.

Sawi, T., Holtzman, B.K., Walter, F., & Paisley, J. (2022). *Journal of Geophysical Research: Earth Surface*



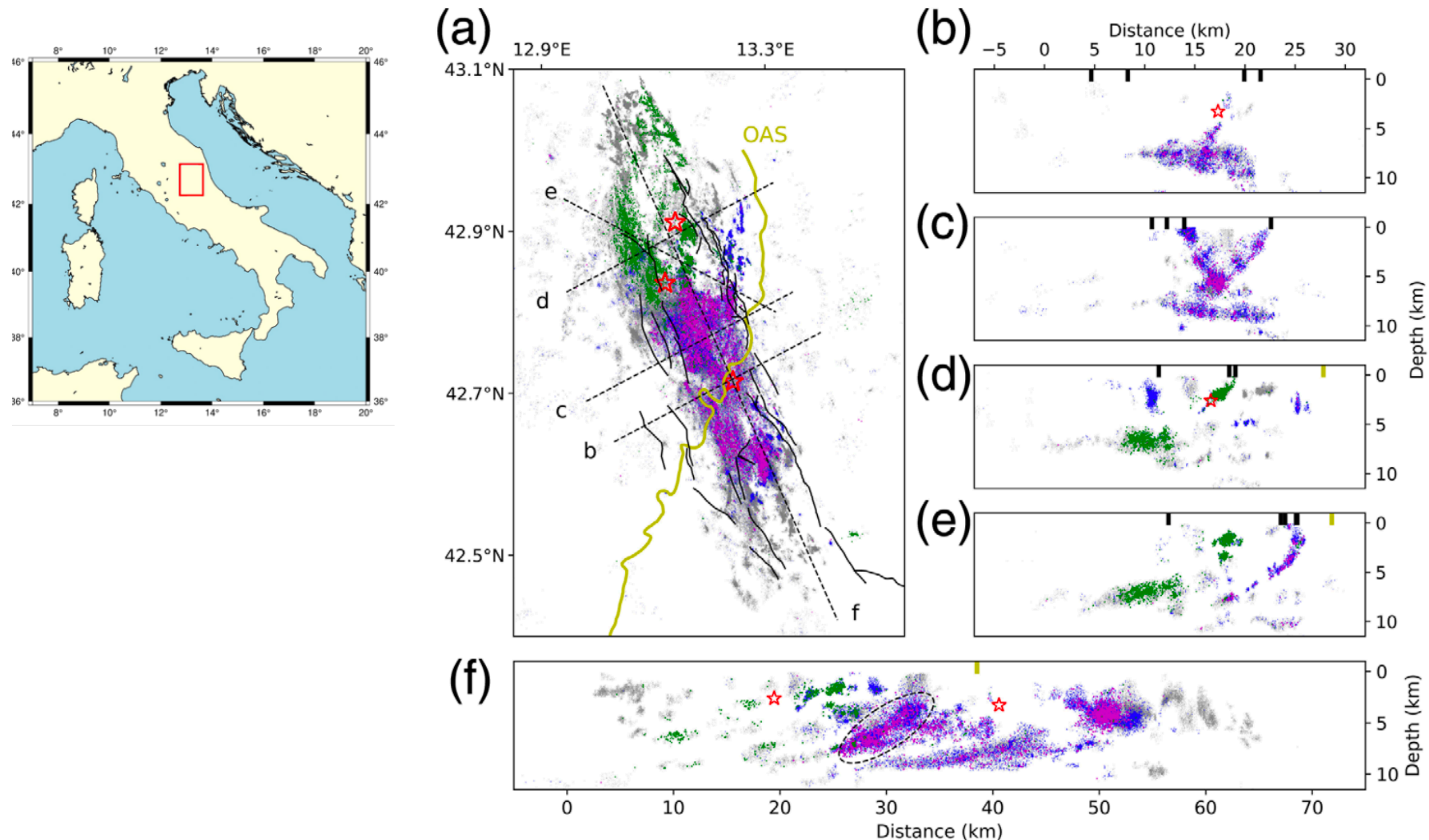
Detecting Repeating Earthquakes on the San Andreas Fault with Unsupervised Machine Learning of Spectrograms (2023) Sawi T.,

Waldhauser F., Holtzman, B.K., Groebner, N.; . *The Seismic Record*



Amatrice-SpecUFEx Tutorial

(subset = 6 days, ~1000 earthquakes)

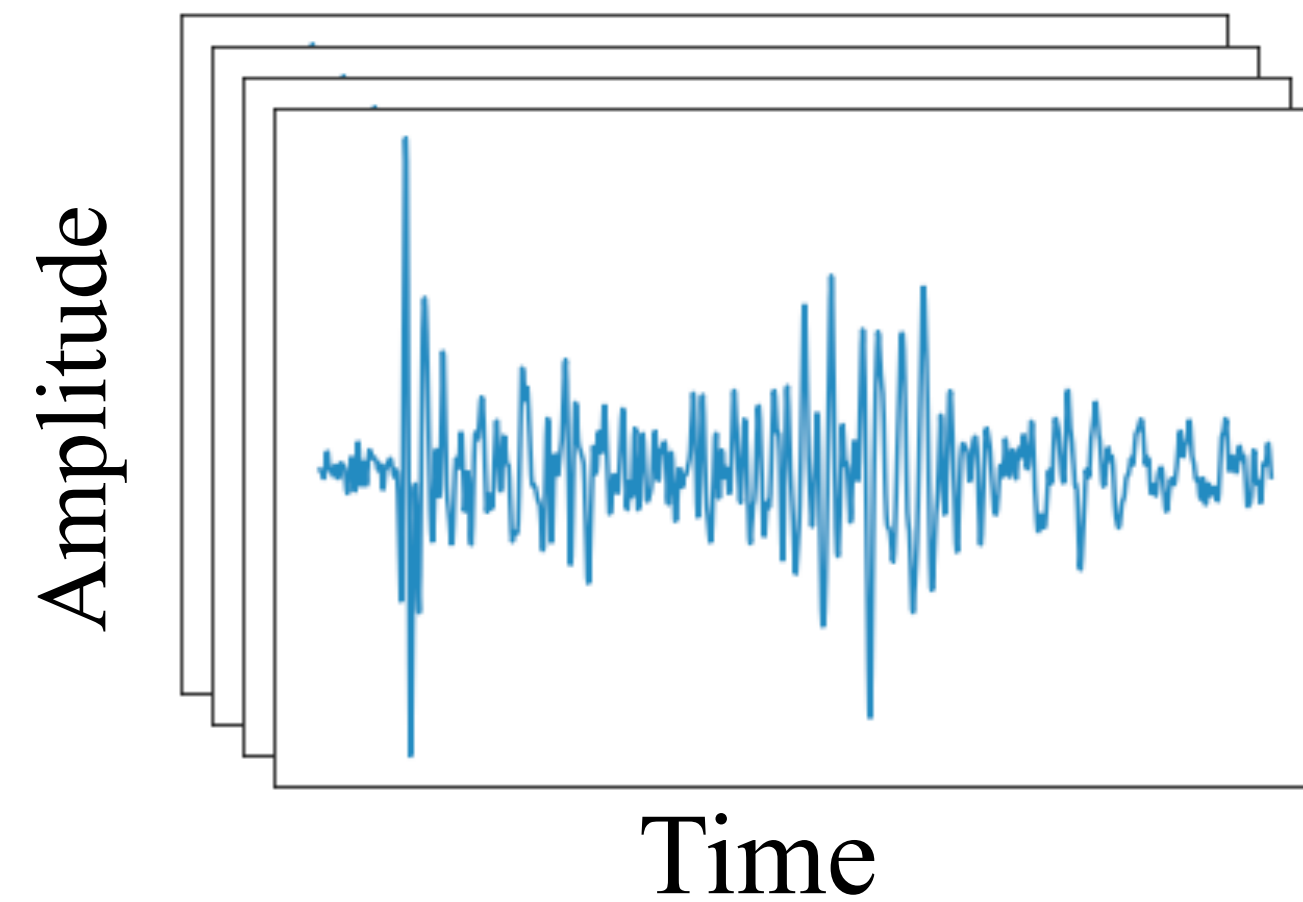


Tan et al., 2021; The Seismic Record

Amatrice-SpecUFEx Tutorial

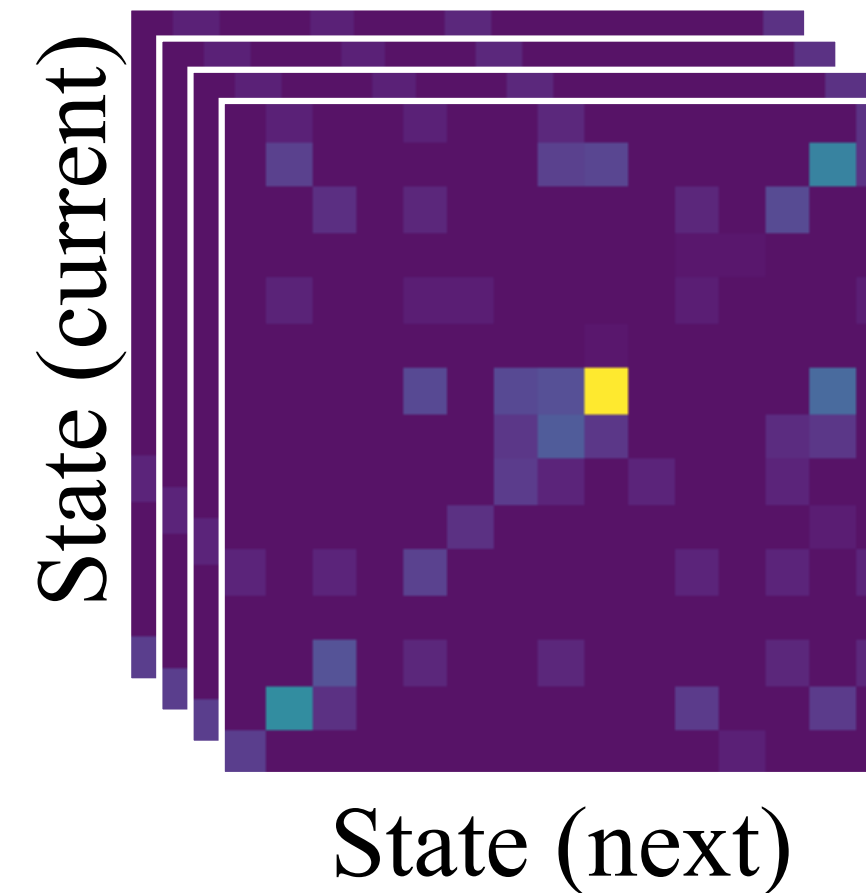
Data:

Earthquake waveforms



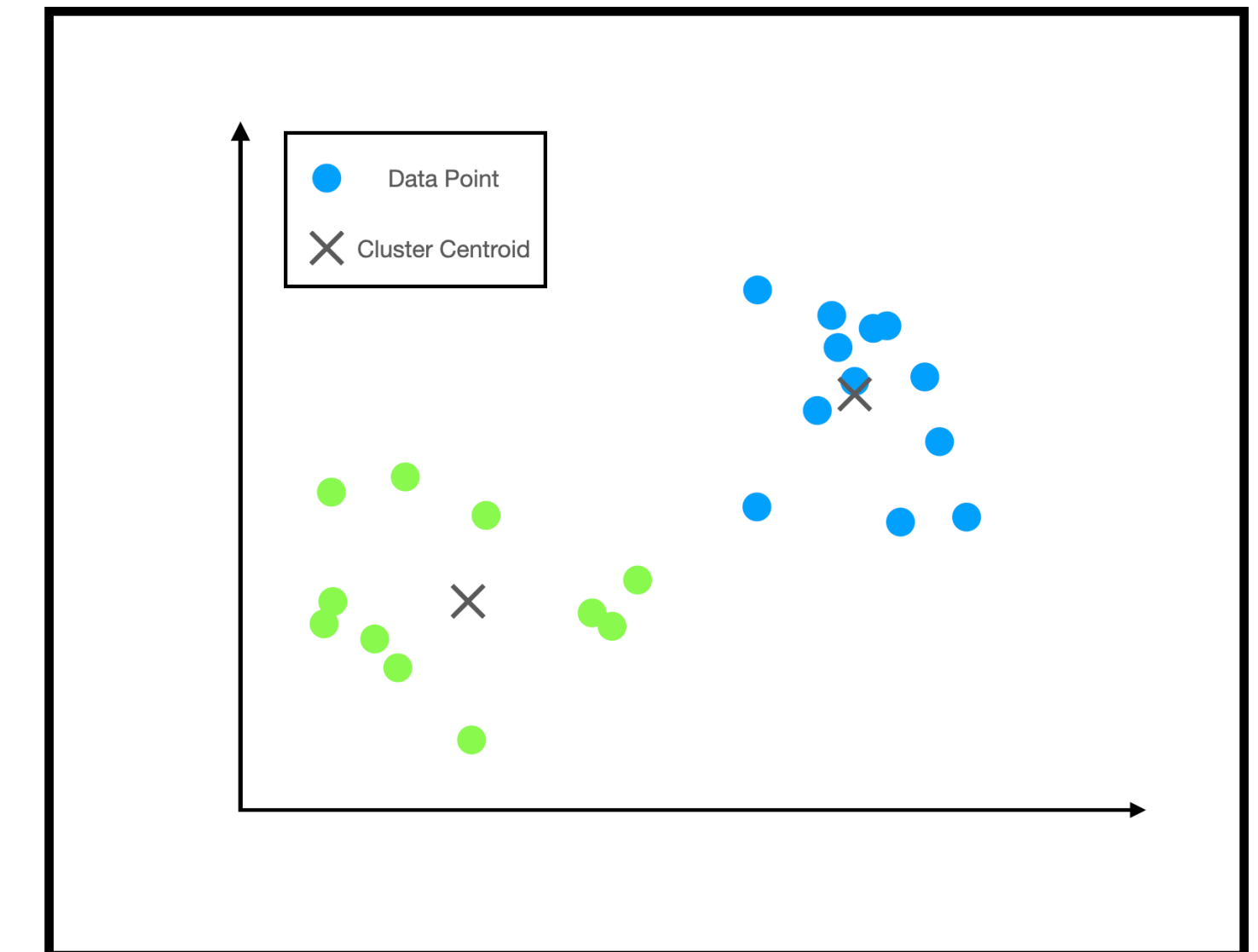
Feature Extraction:

NMF and hidden Markov model



Clustering:

K-means with silhouette scores



Amatrice-SpecUFEx Tutorial

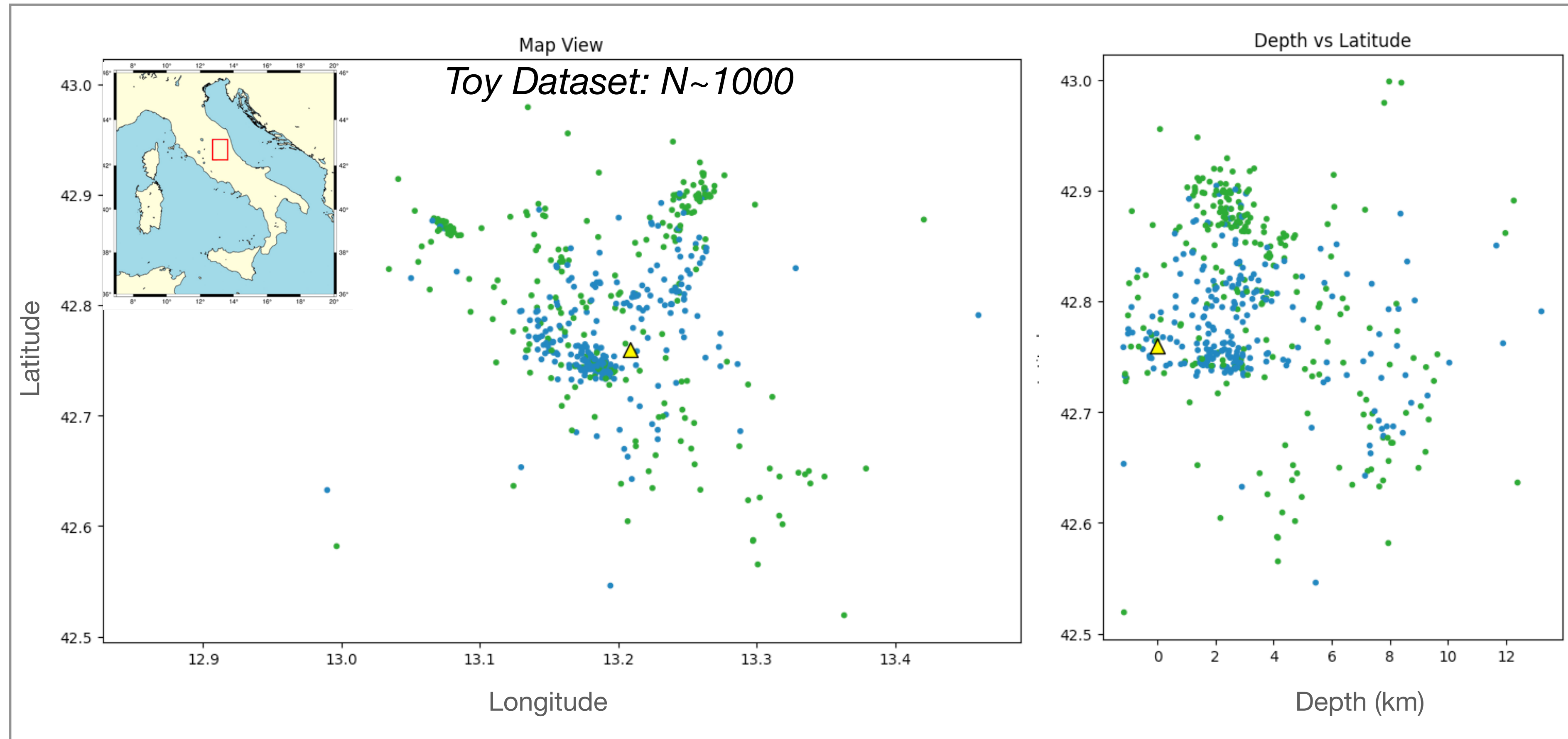
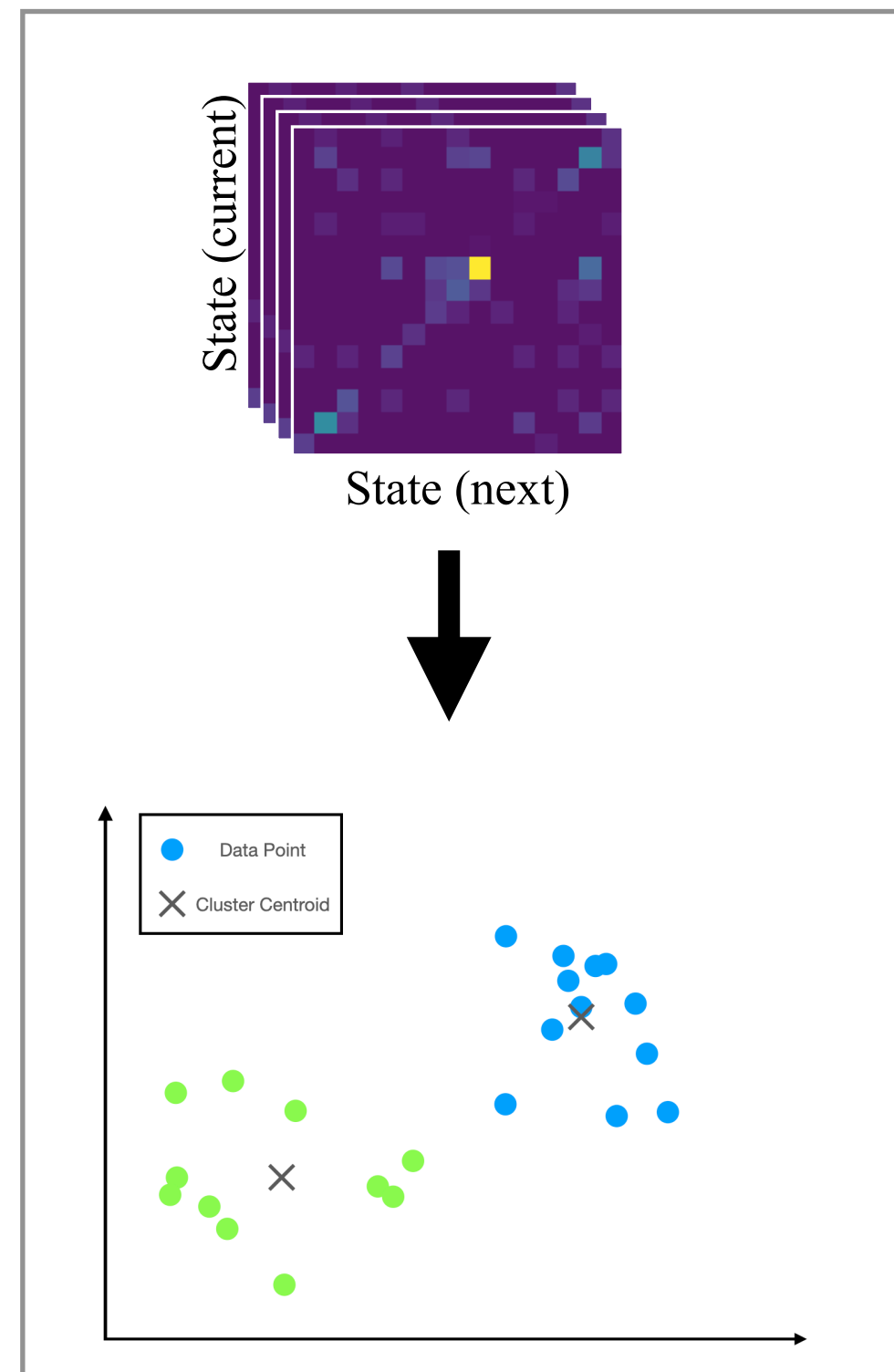
SpecUFEx

+

K-means clustering

Results:

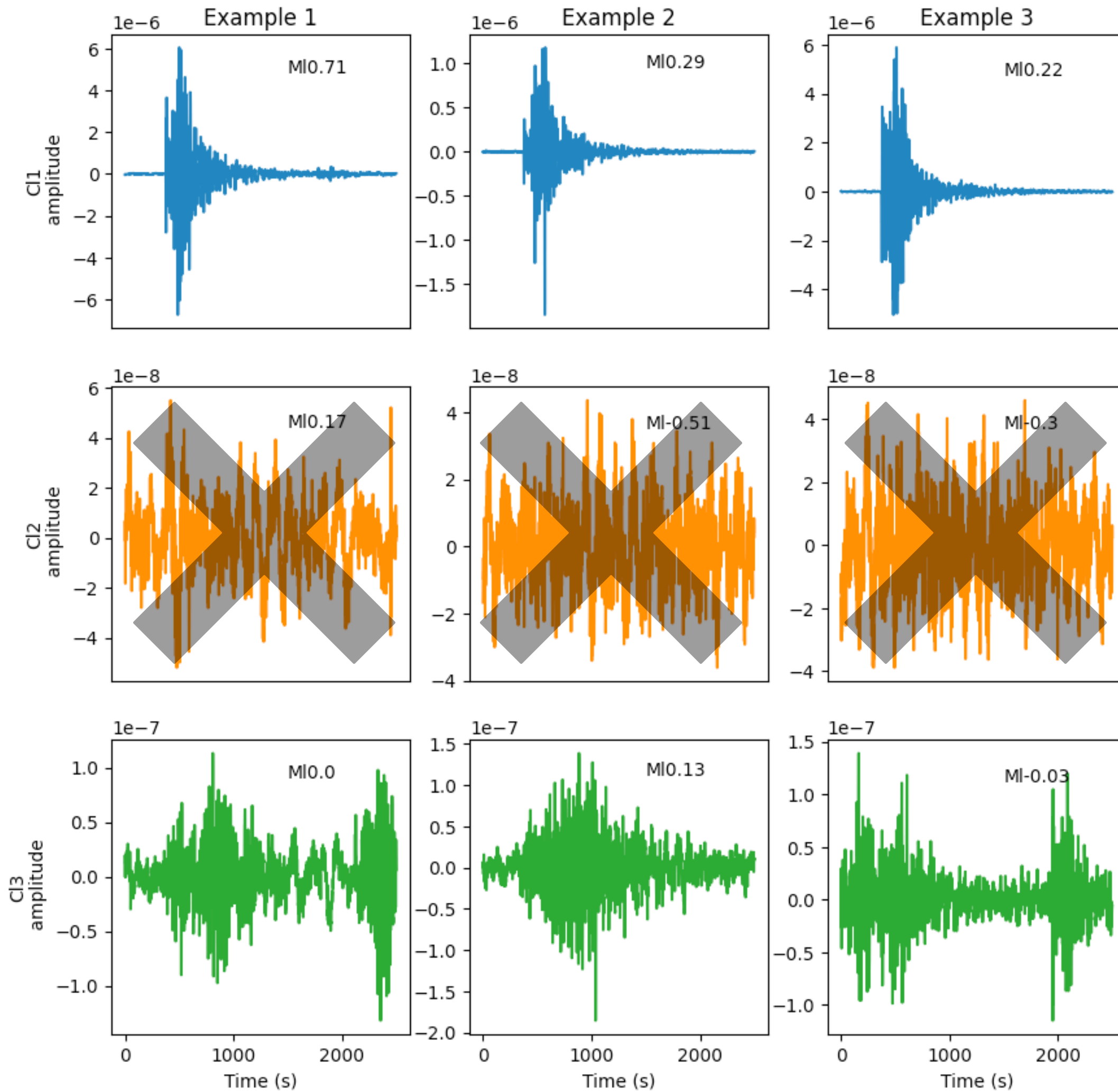
Map showing cluster assignments



Results for $K=3$, with 2nd cluster removed

Amatrice-SpecUEx Tutorial

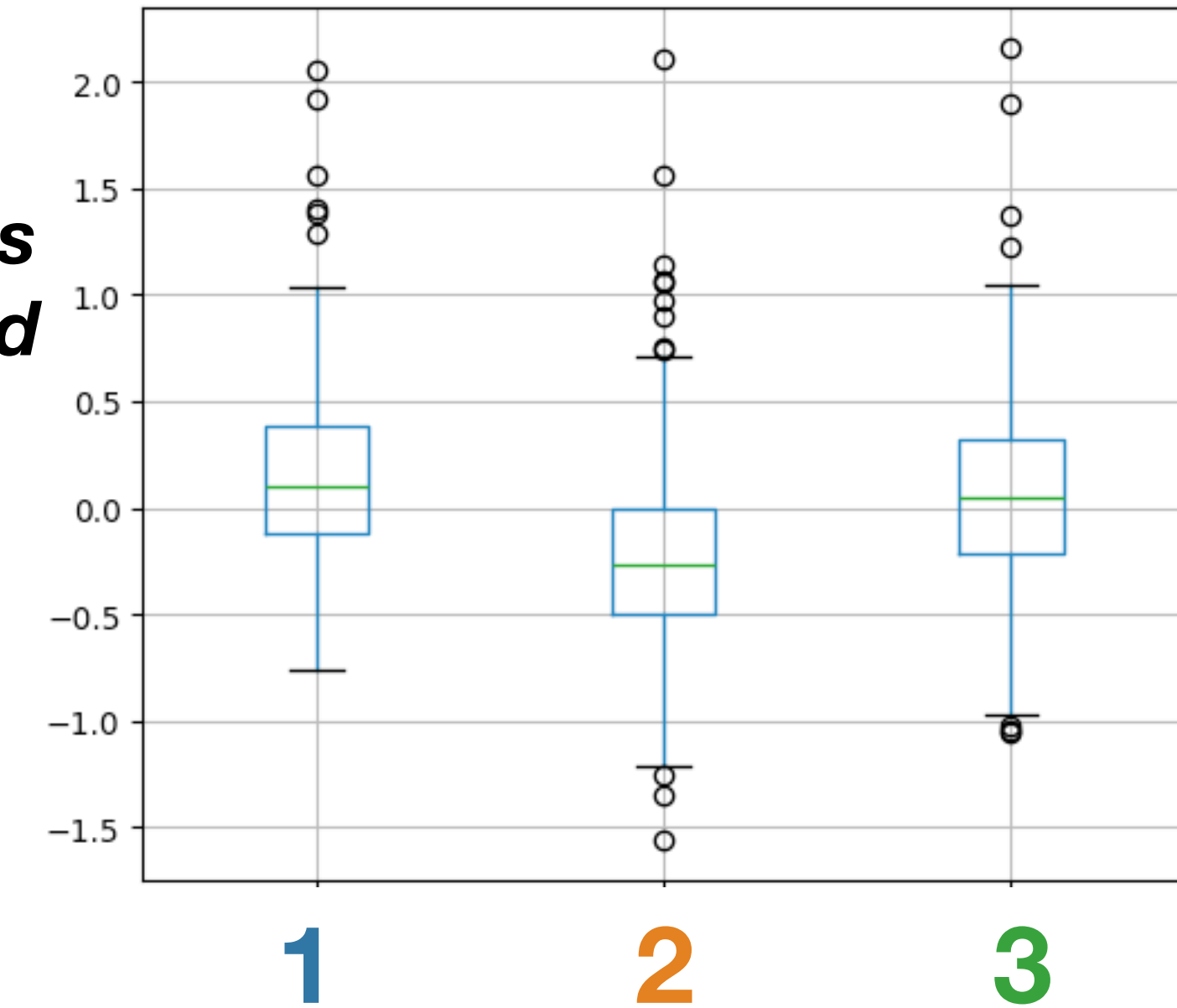
**Waveform cluster examples -
Varying frequency content,
duration, SNR**



**Noisy waveforms
- bad picks?**

**Similar
magnitudes
for blue and
green**

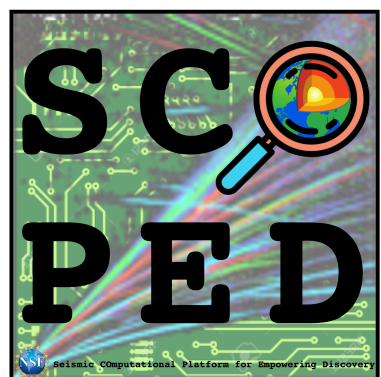
Estimated magnitudes



Summary:

- **Identifying different families of waveforms**
 - (no templates or CC used)
- **Spatial relationship; distinct waveforms; not magnitude-dependent**
- **Different mechanisms/fault conditions?**
- **To Do: Compare other stations**

Thank you! Questions?



Lamont-Doherty Earth Observatory
COLUMBIA UNIVERSITY | EARTH INSTITUTE



Unsupervised Machine Learning for Seismology Using AWS

SCOPED University of Washington, Seattle 2024

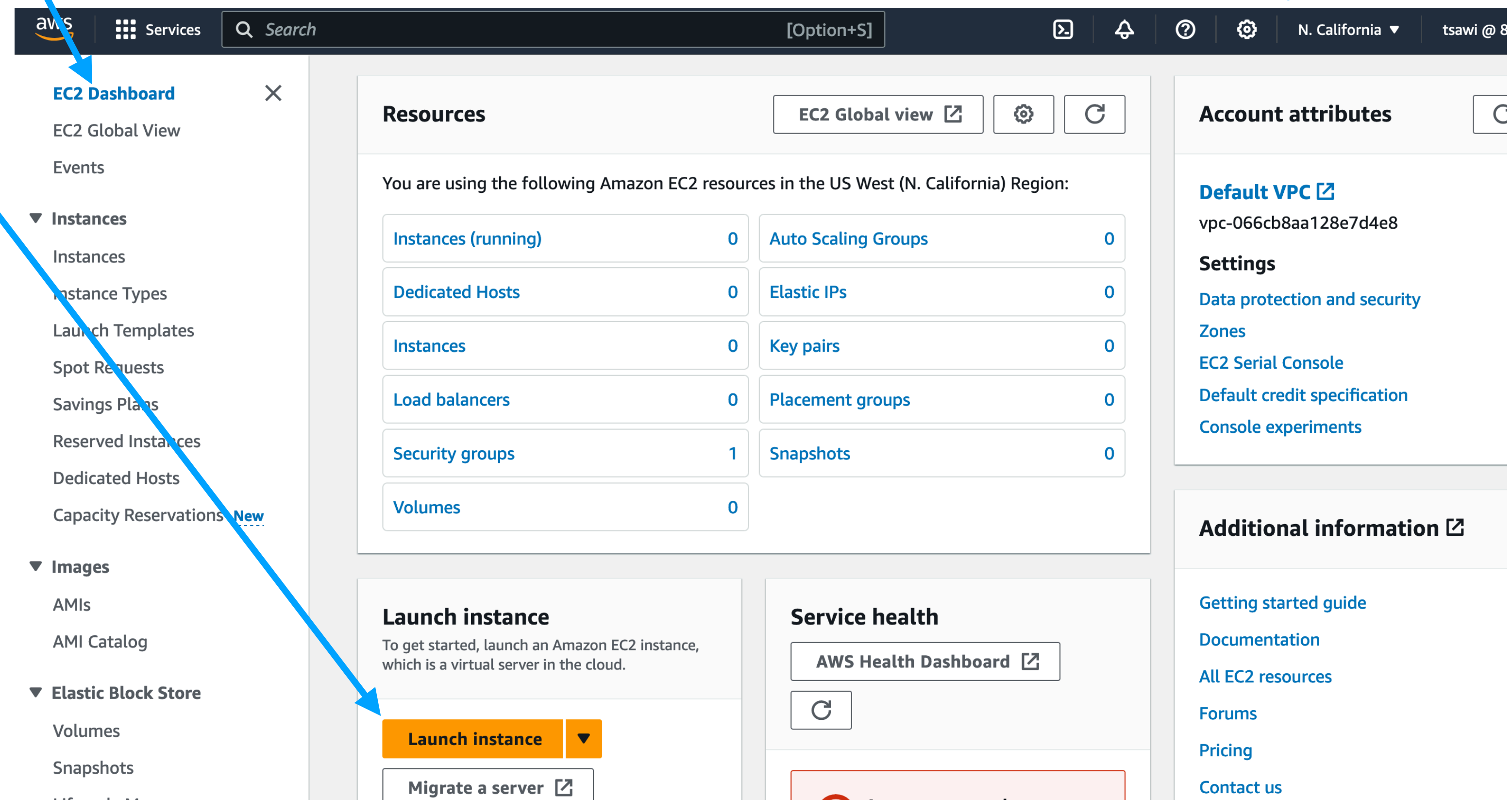
Theresa Sawi, Nate Groebner, Felix Waldhauser, Kaiwen Wang, Yen-Joe Tan

Thursday 5/23/2024

1:00 - 1:45	Lecture on unsupervised machine learning in seismology
1:45 - 2:00	Q&A
2:00 - 2:10	Break now
2:10 - 3:00	Tutorial on SpecUFEx/2016 Amatrice sequence

Using AWS Cloud

1. Sign in to AWS console (<https://aws.amazon.com/>)
2. Make sure you are in West Coast region (us-west-1)
3. Navigate to Dashboard
4. Launch instance



Using AWS Cloud

1. **Name:** your choosing
2. **Instance type:** t2: large
3. **Network settings:** Select existing security group (“default”)
4. **Launch instance**

The screenshot shows the 'Launch instance' page in the AWS Management Console. Blue arrows from the instructions point to the following elements:

- Arrow 1 points to the 'Name' field in the 'Summary' section.
- Arrow 2 points to the 'Instance type' dropdown, which is set to 't2.large'.
- Arrow 3 points to the 'Select existing security group' radio button in the 'Firewall (security groups)' section.
- Arrow 4 points to the 'Launch instance' button at the bottom right.

Network settings [Info](#) [Edit](#)

Network [Info](#)
vpc-066cb8aa128e7d4e8

Subnet [Info](#)
No preference (Default subnet in any availability zone)

Auto-assign public IP [Info](#)
Enable

Additional charges apply when outside of free tier allowance

Firewall (security groups) [Info](#)
A security group is a set of firewall rules that control the traffic for your instance. Add rules to allow specific traffic to reach your instance.

☐ Create security group ☒ Select existing security group

Common security groups [Info](#)
Select security groups

default sg-0f987439f9c180fde X
VPC: vpc-066cb8aa128e7d4e8

[Compare security group rules](#)

Security groups that you add or remove here will be added to or removed from all your network interfaces.

Configure storage [Info](#) [Advanced](#)

1x 8 GiB gp3 Root volume (Not encrypted)

Summary

Number of instances [Info](#)
1

Software Image (AMI)
Amazon Linux 2023 AMI 2023.4.2...[read more](#)
ami-0cbe318e714fc9a82

Virtual server type (instance type)
t2.large

Firewall (security group)
default

Storage (volumes)
1 volume(s) - 8 GiB

Free tier: In your first year includes 750 hours of t2.micro (or t3.micro in the Regions in which t2.micro is unavailable) instance usage on free tier AMIs per month, 750 hours of public IPv4 address usage per month, 30 GiB of EBS storage, 2 million IOs, 1 GB of snapshots, and

[Cancel](#) [Launch instance](#) [Review commands](#)

Using AWS Cloud

1. Select “proceed without key pair”
2. Launch instance

Select an existing key pair or create a key pair

Information We noticed that you didn't select a key pair. If you want to be able to connect to your instance it is recommended that you create one or select an existing one.

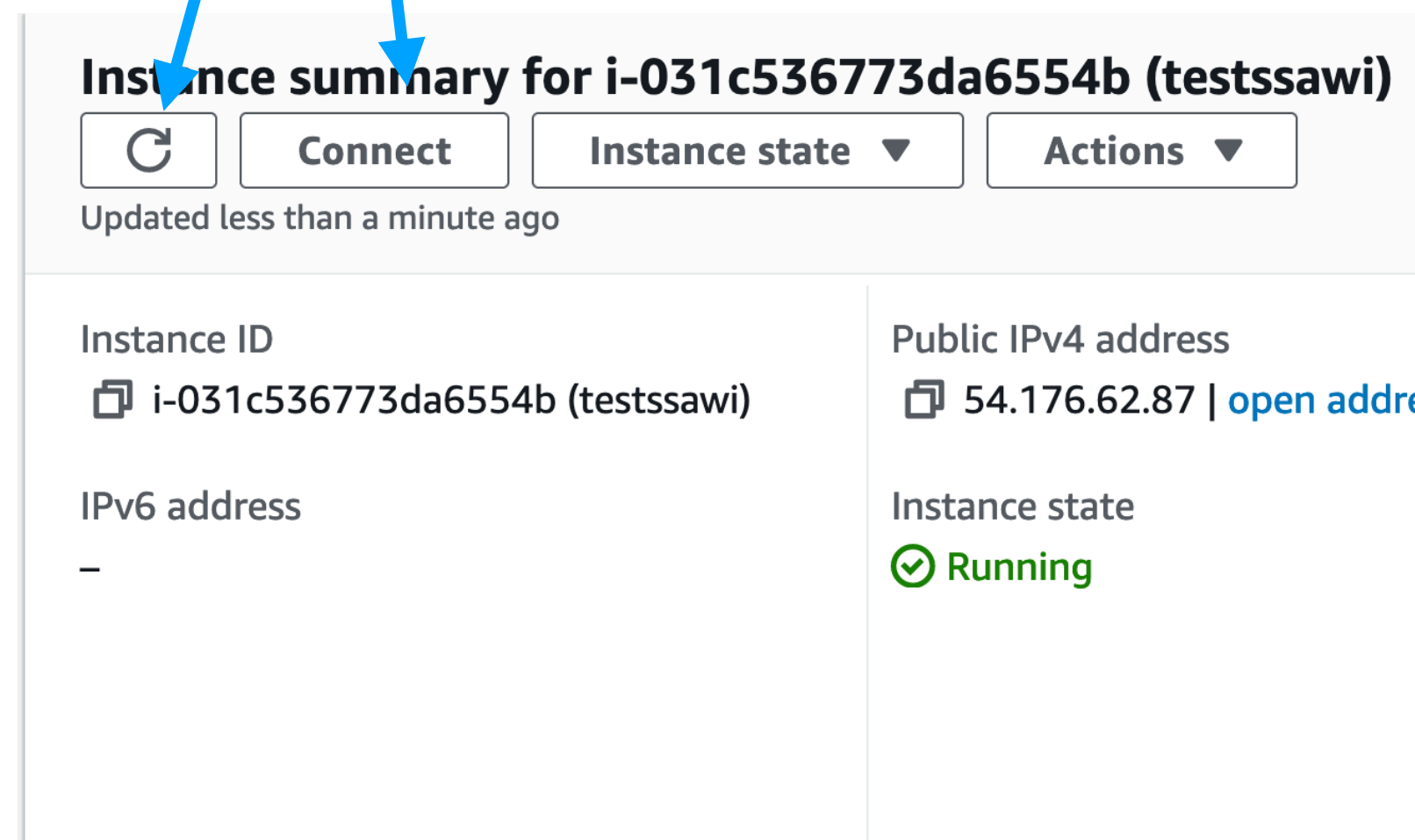
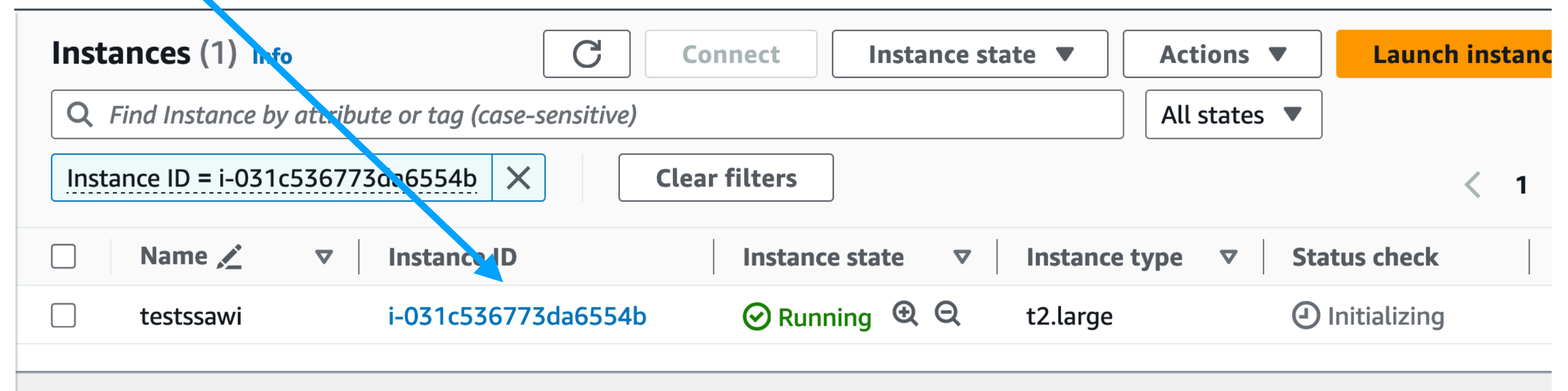
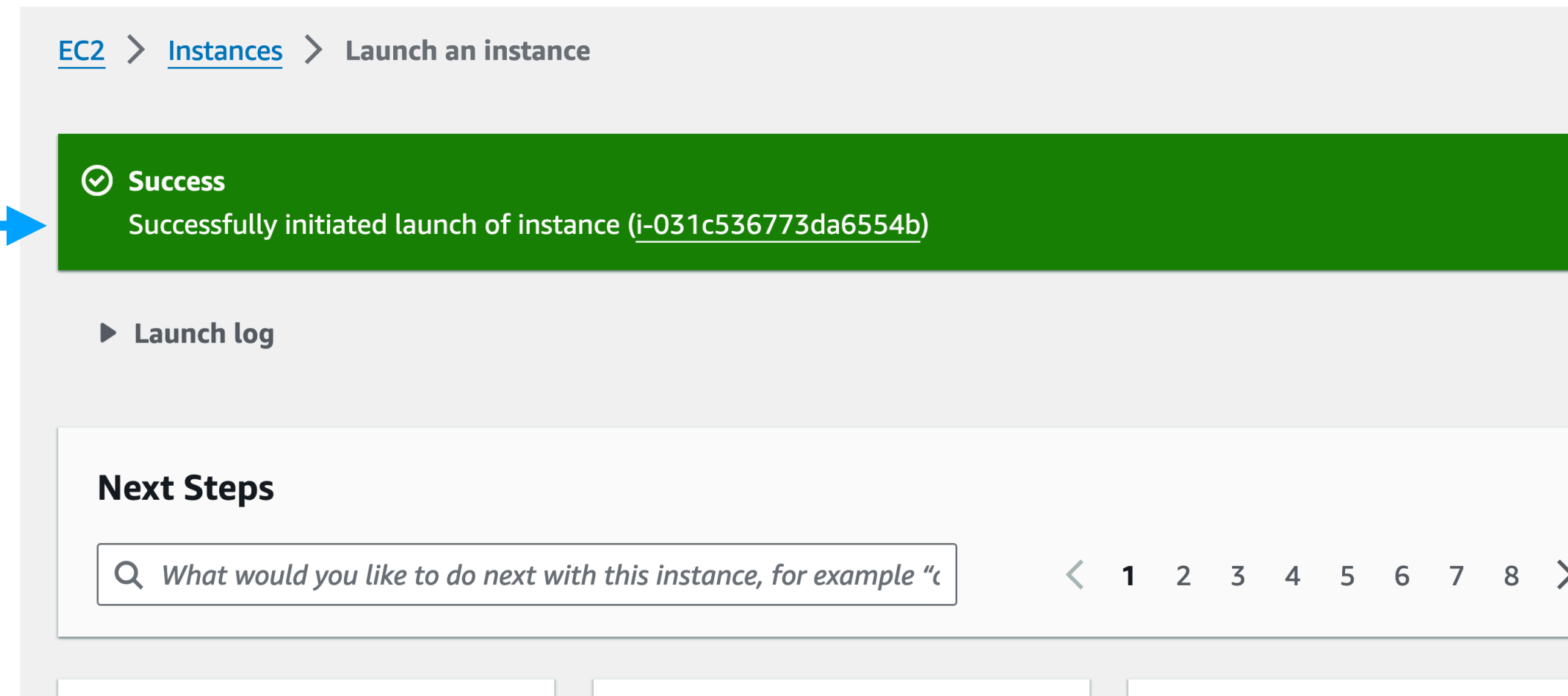
☐ Create new key pair ☒ Proceed without key pair

Information Note that you will not be able to connect to this instance unless you already know the password built into this AMI.

Cancel Launch instance

Using AWS Cloud

1. Click instance link
2. Click instance link again
3. Click “connect”, which you may need to refresh a couple times



Using AWS Cloud

1. Under the EC2 Instance Connect tab, click “Connect”

This should open the linux terminal with a blank machine:

```

#
' \ ##### Amazon Linux 2023
~~ \ ##### \
~~ \ ### |
~~ \ #/ https://aws.amazon.com/linux/amazon-linux-2023
~~ V ~ ' ' ->
~~
~~ . .
~~ \ / \ / \ /
~~ /m/ ' ' \ /
ec2-user@ip-172-31-23-105 ~]$ sudo yum install -y git docker
yget https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.sh
chmod +x Miniconda3-latest-Linux-x86_64.sh
./Miniconda3-latest-Linux-x86_64.sh -b -p $HOME/miniconda
./miniconda/bin/conda init bash
bash

```

EC2 Instance Connect

Session Manager

SSH client

EC2 serial console

**Port 22 (SSH) is not authorized**
Port 22 (SSH) is currently not authorized by your security group. To use EC2 Instance Connect, you must authorize port 22 for the EC2 Instance Connect service IP addresses in your Region: 13.52.6.112/29. [Learn more.](#)

Instance ID

 i-031c536773da6554b (testssawi)

Connection Type

☒ **Connect using EC2 Instance Connect**
Connect using the EC2 Instance Connect browser-based client, with a public IPv4 address.

☐ **Connect using EC2 Instance Connect Endpoint**
Connect using the EC2 Instance Connect browser-based client, with a private IPv4 address and a VPC endpoint.

Public IP address

 54.176.62.87

Username

Enter the username defined in the AMI used to launch the instance. If you didn't define a custom username, use the default username, ec2-user.

 ec2-user



 **Note:** In most cases, the default username, ec2-user, is correct. However, read your AMI usage instructions to check if the AMI owner has changed the default AMI username.

Cancel

Connect

Tutorial Notebook (static)

sawilabs.com/cloud.html

These Slides

sawilabs.com/slides.pdf

Download Docker image locally

<https://github.com/SeisSCOPED/specufex>