

Lamont-Doherty Earth Observatory
COLUMBIA UNIVERSITY | EARTH INSTITUTE



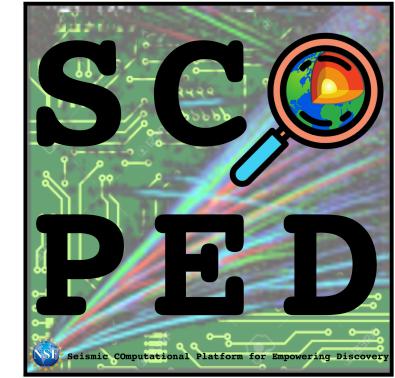
Unsupervised Machine Learning for Seismology Using AWS

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**

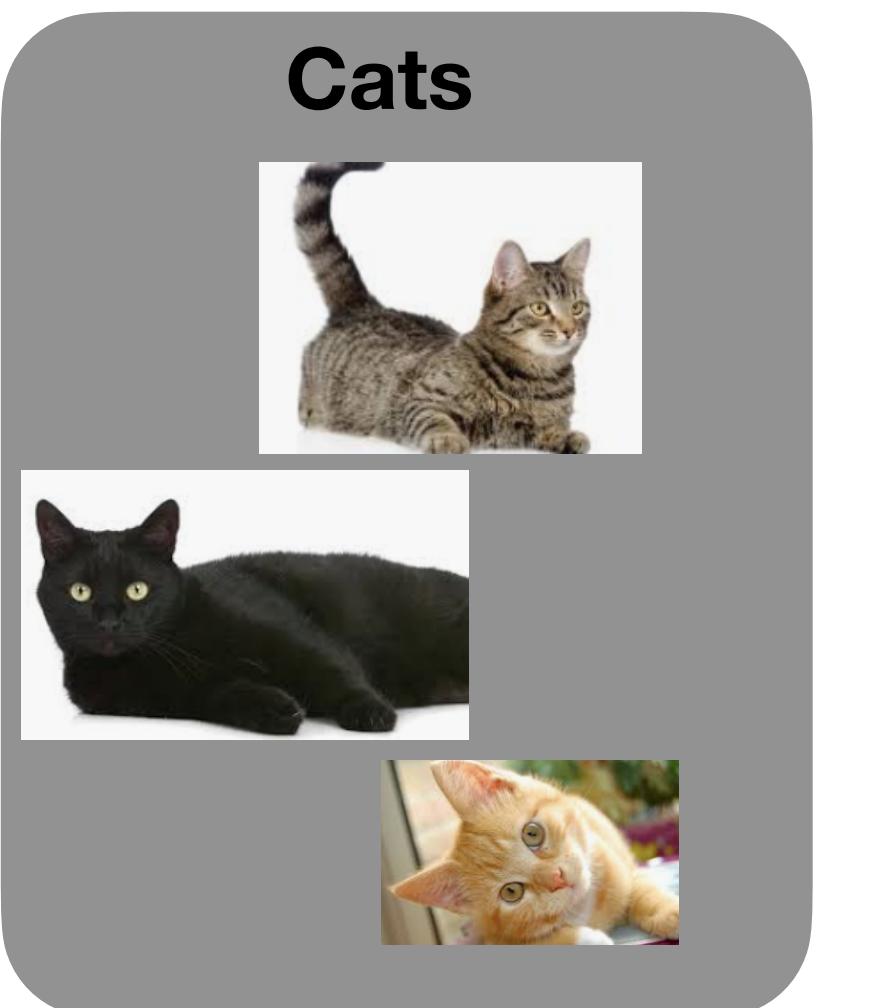
Supervised Machine Learning

Labeled (training) dataset



Supervised Machine Learning

Labeled (training) dataset

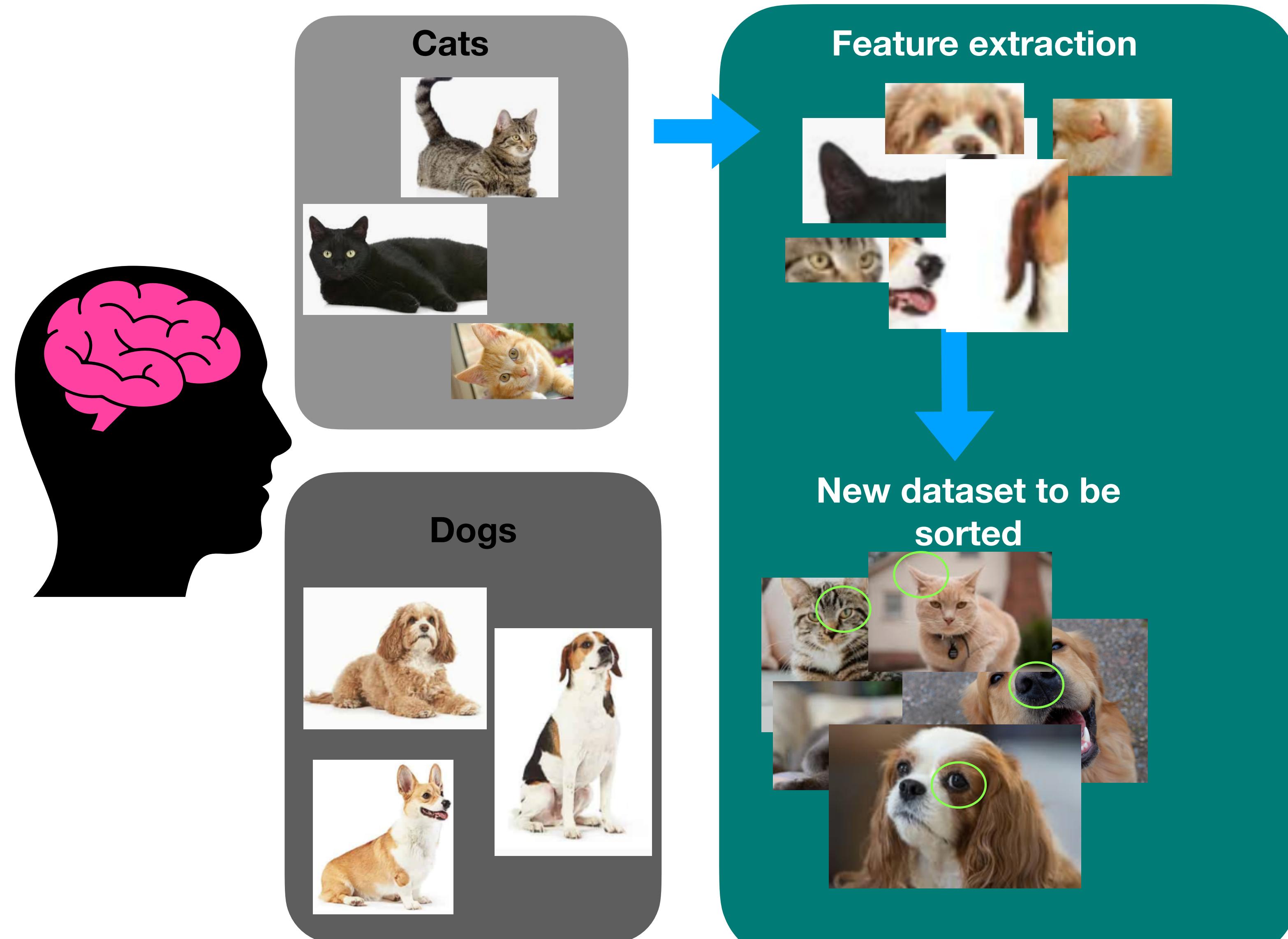


New dataset to be sorted



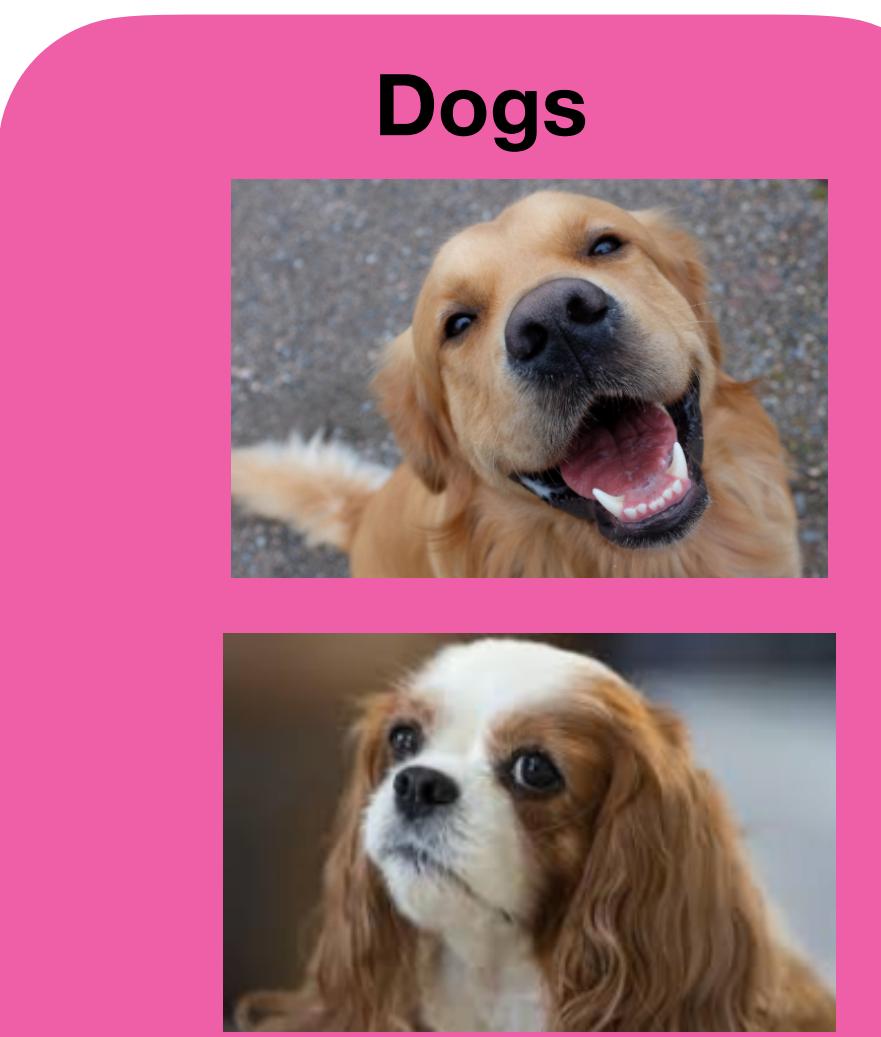
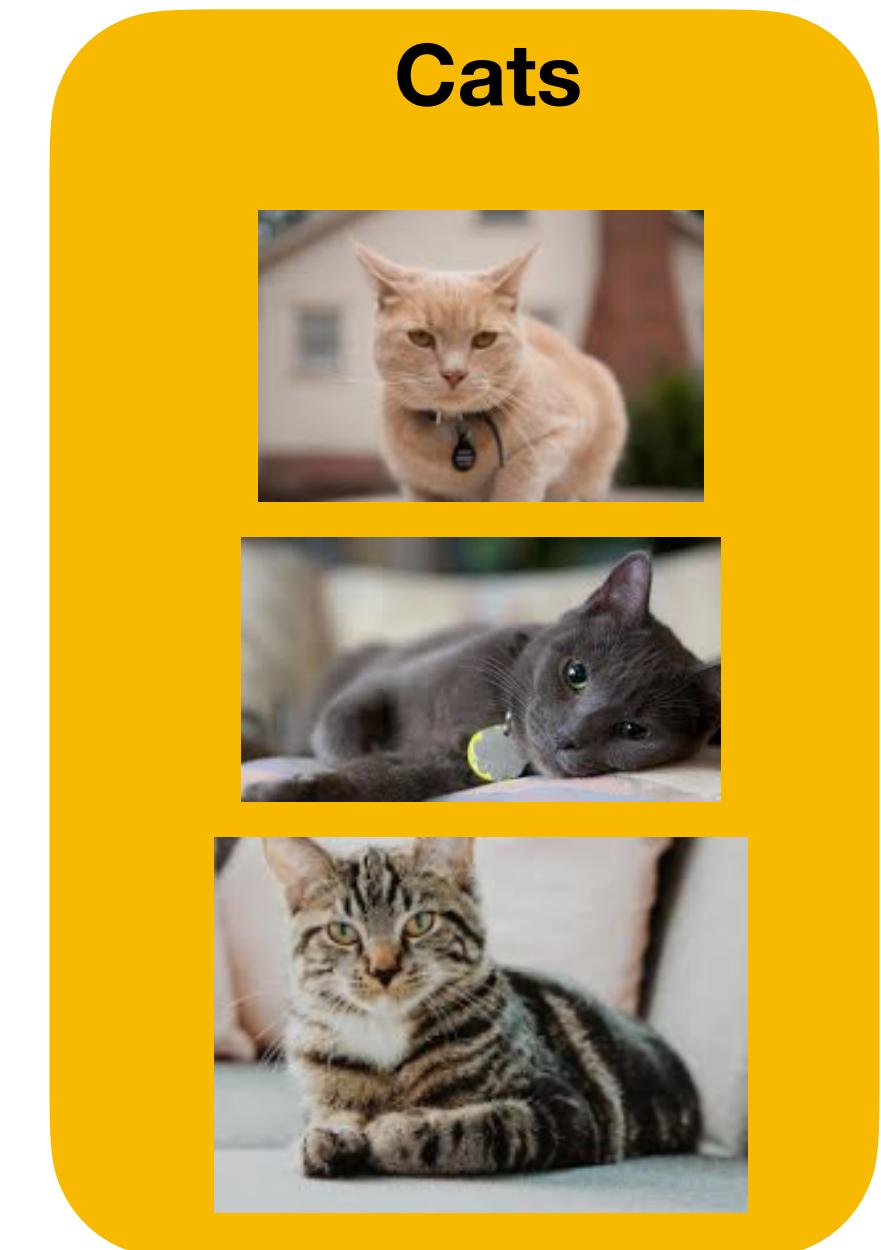
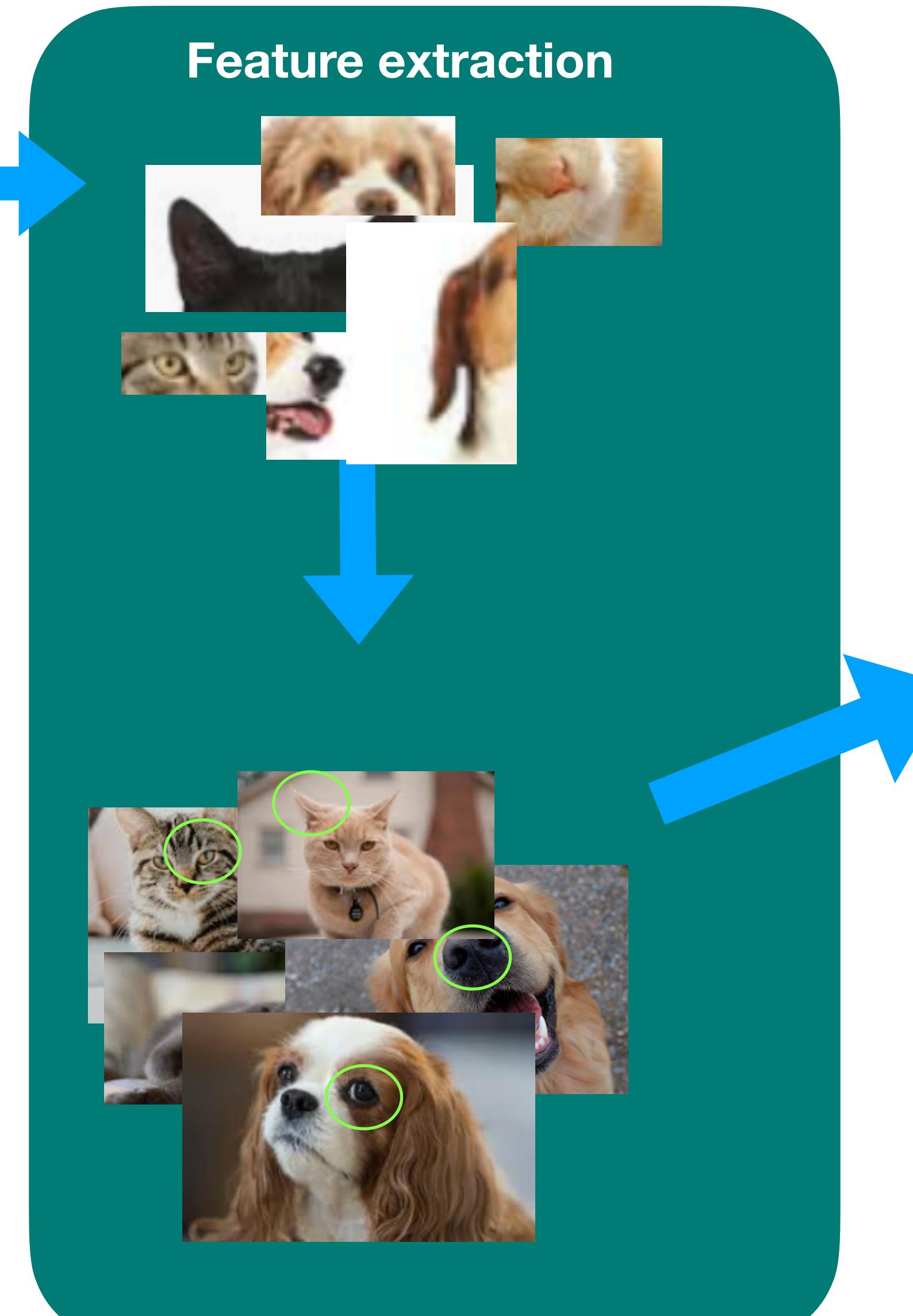
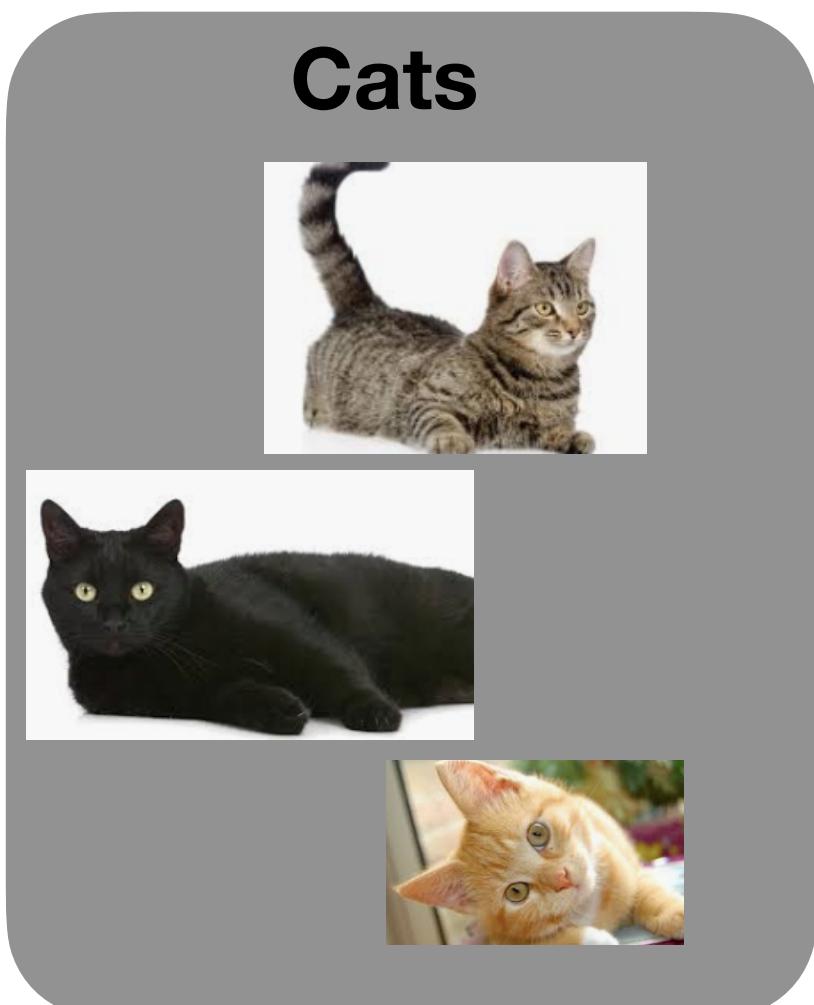
Supervised Machine Learning

Labeled (training) dataset



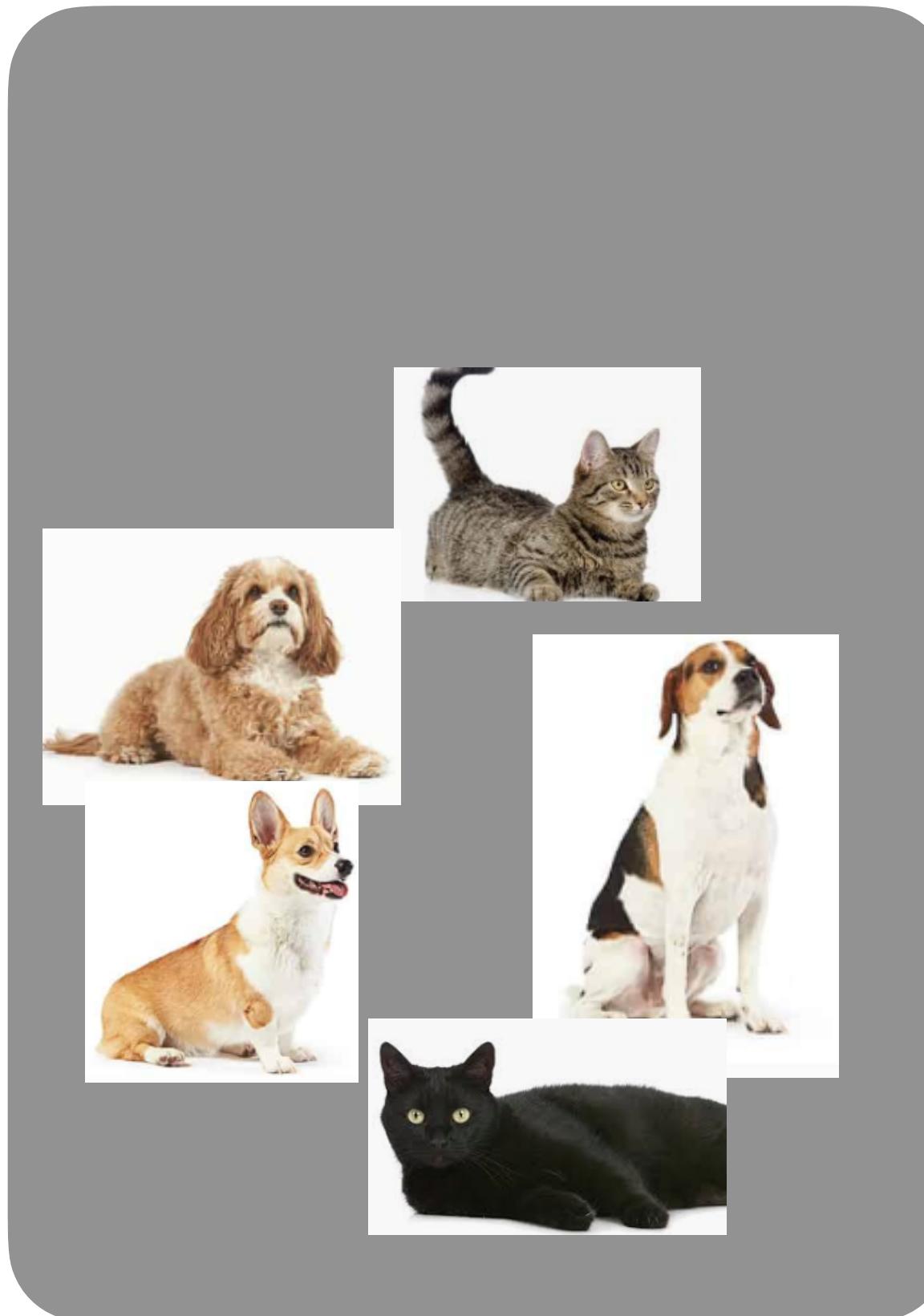
Supervised Machine Learning

Labeled (training) dataset

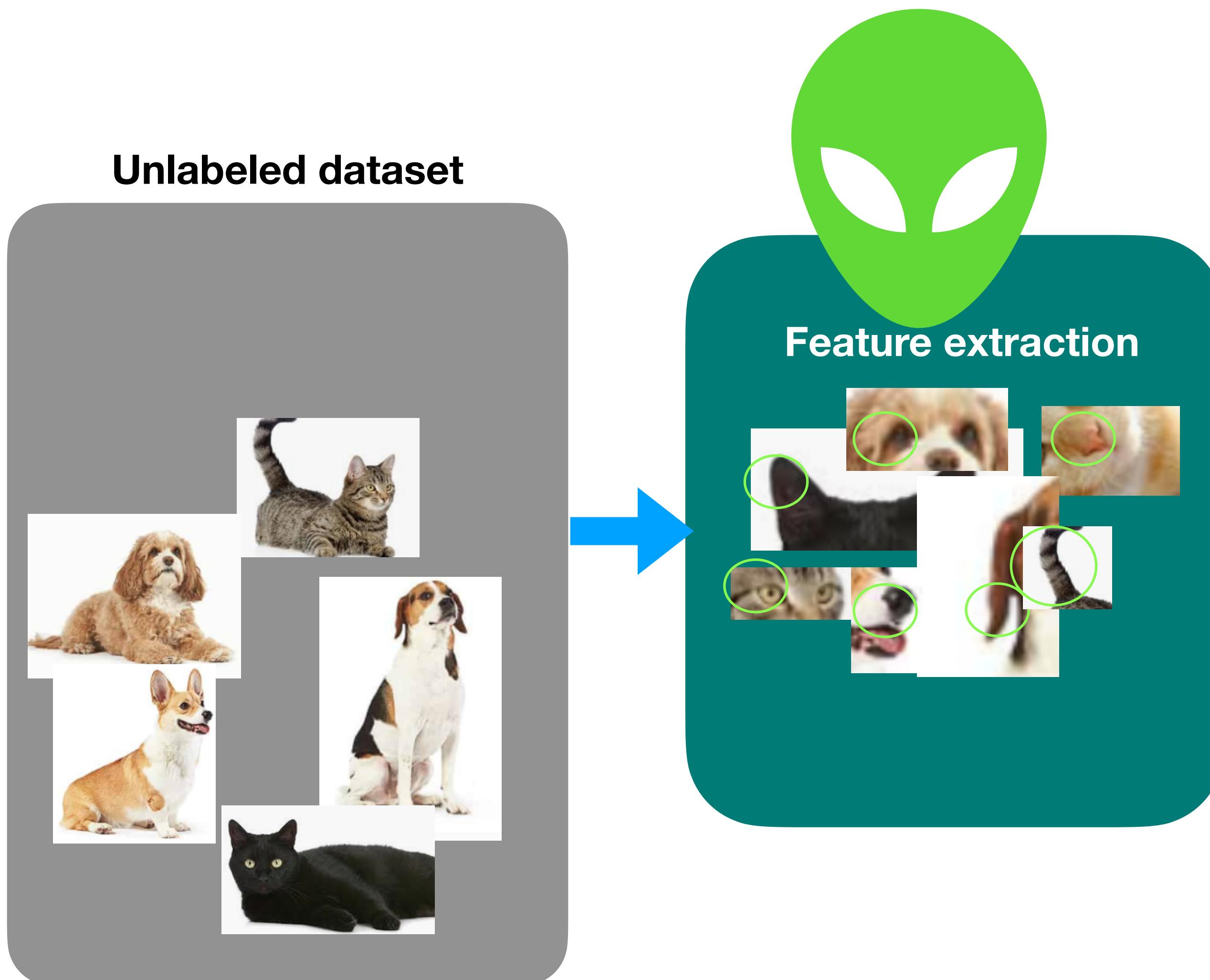


Unsupervised Machine Learning

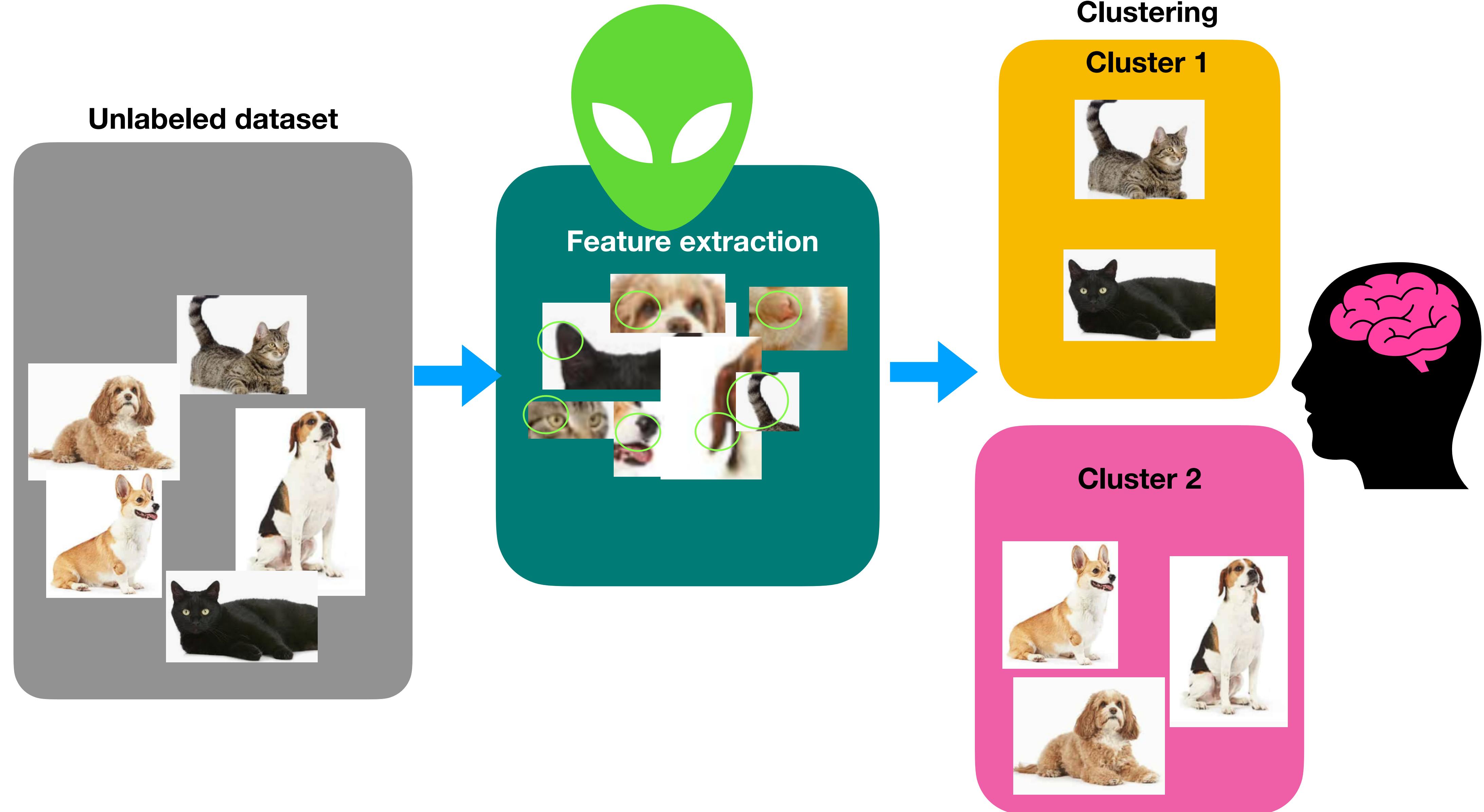
Unlabeled dataset



Unsupervised Machine Learning



Unsupervised Machine Learning



Machine Learning

Supervised

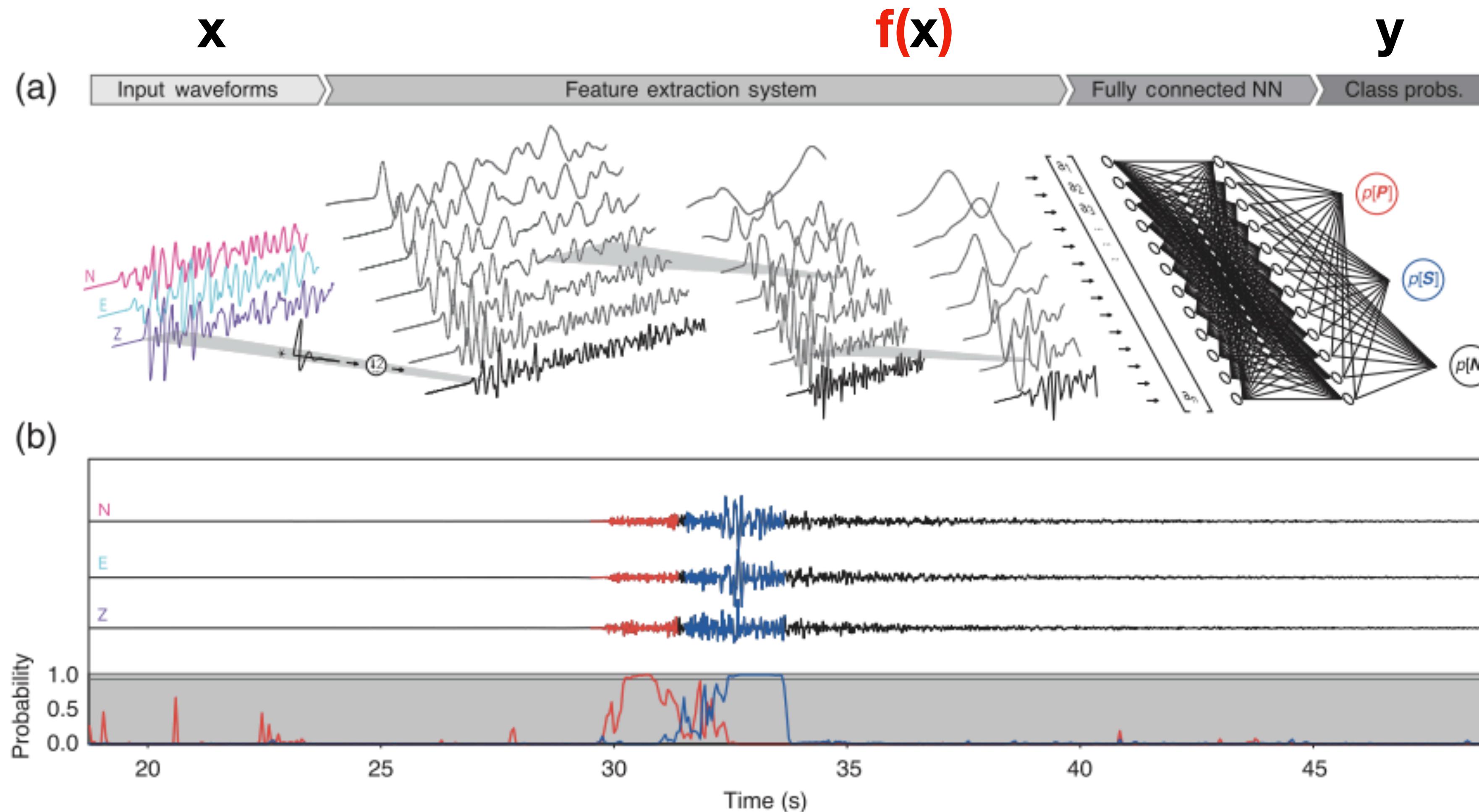
Unsupervised

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

Unsupervised

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

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

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

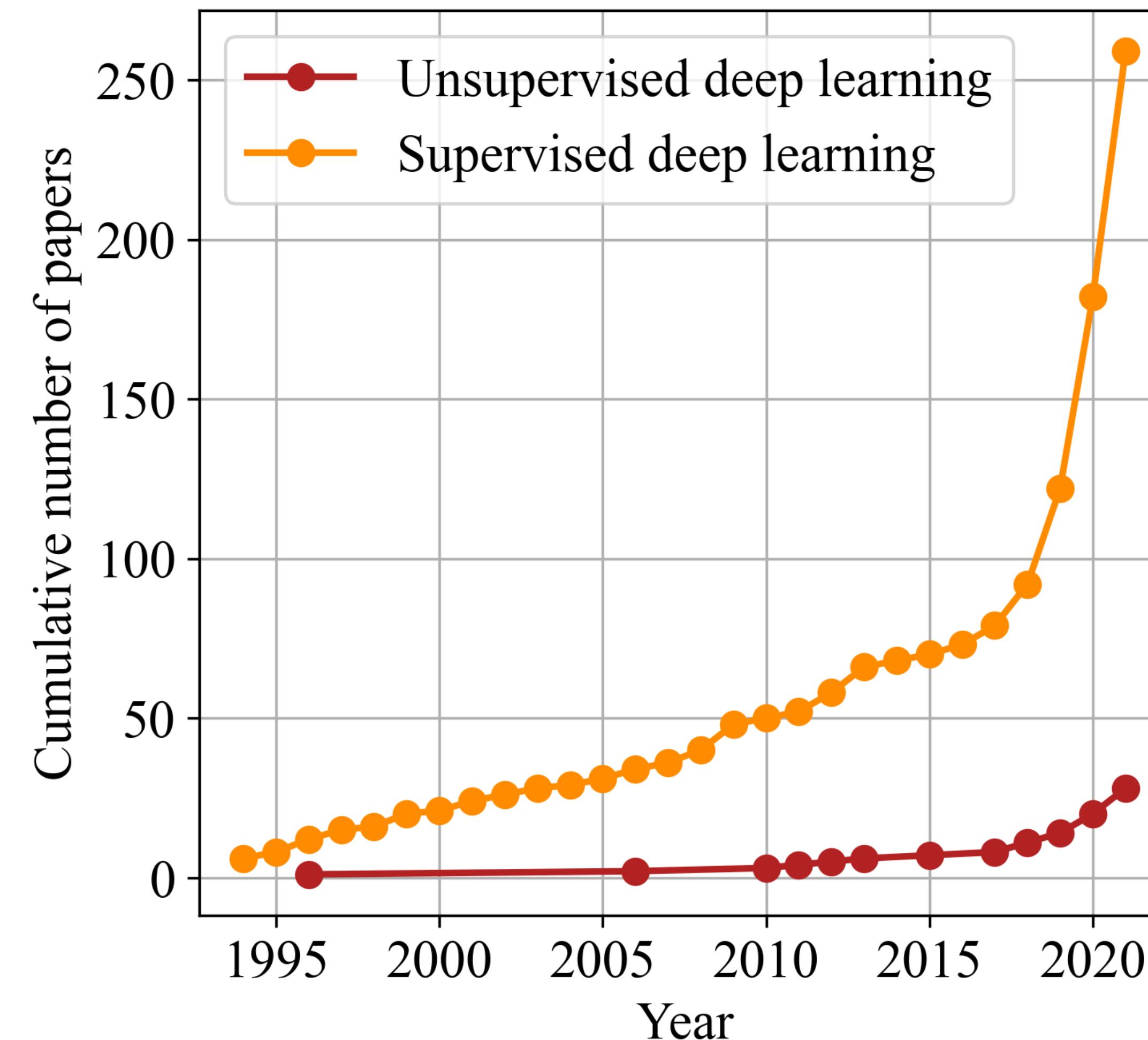
Unsupervised

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**

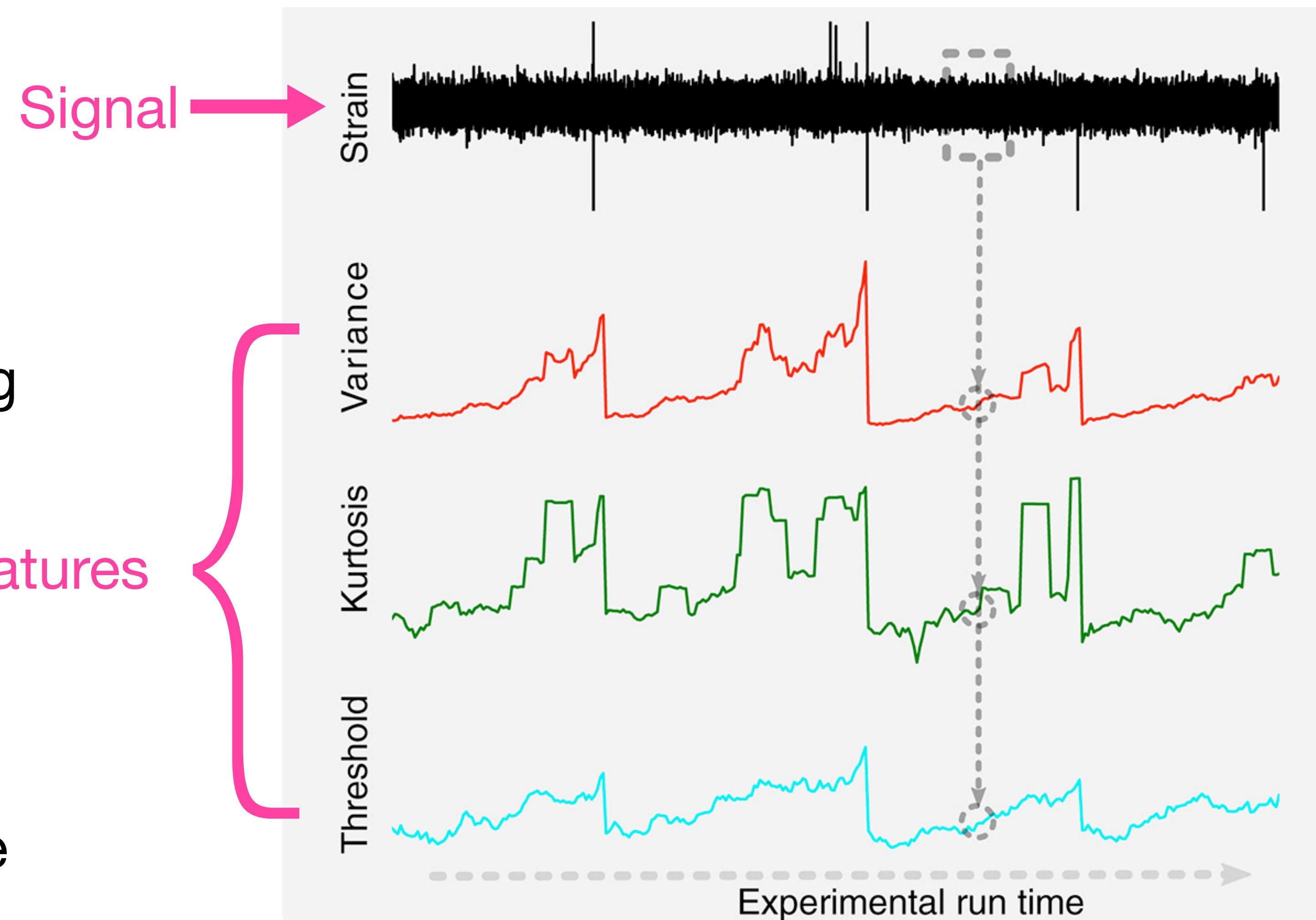
Feature Extraction & Dimensionality Reduction

“Garbage in, garbage out”

Extracting information from data

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

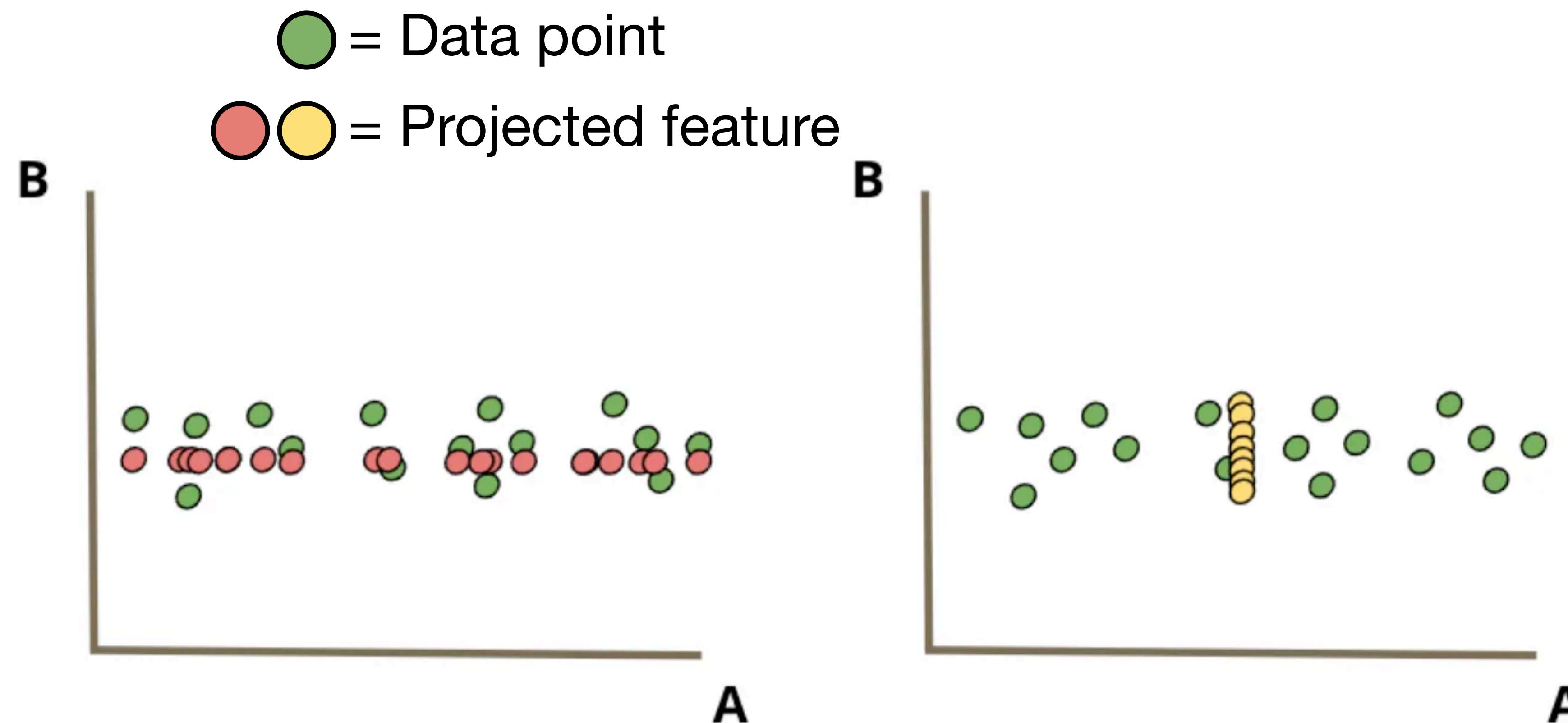
Techniques depend on data type
and domain knowledge



Rouet-Leduc et al., (2017) GRL

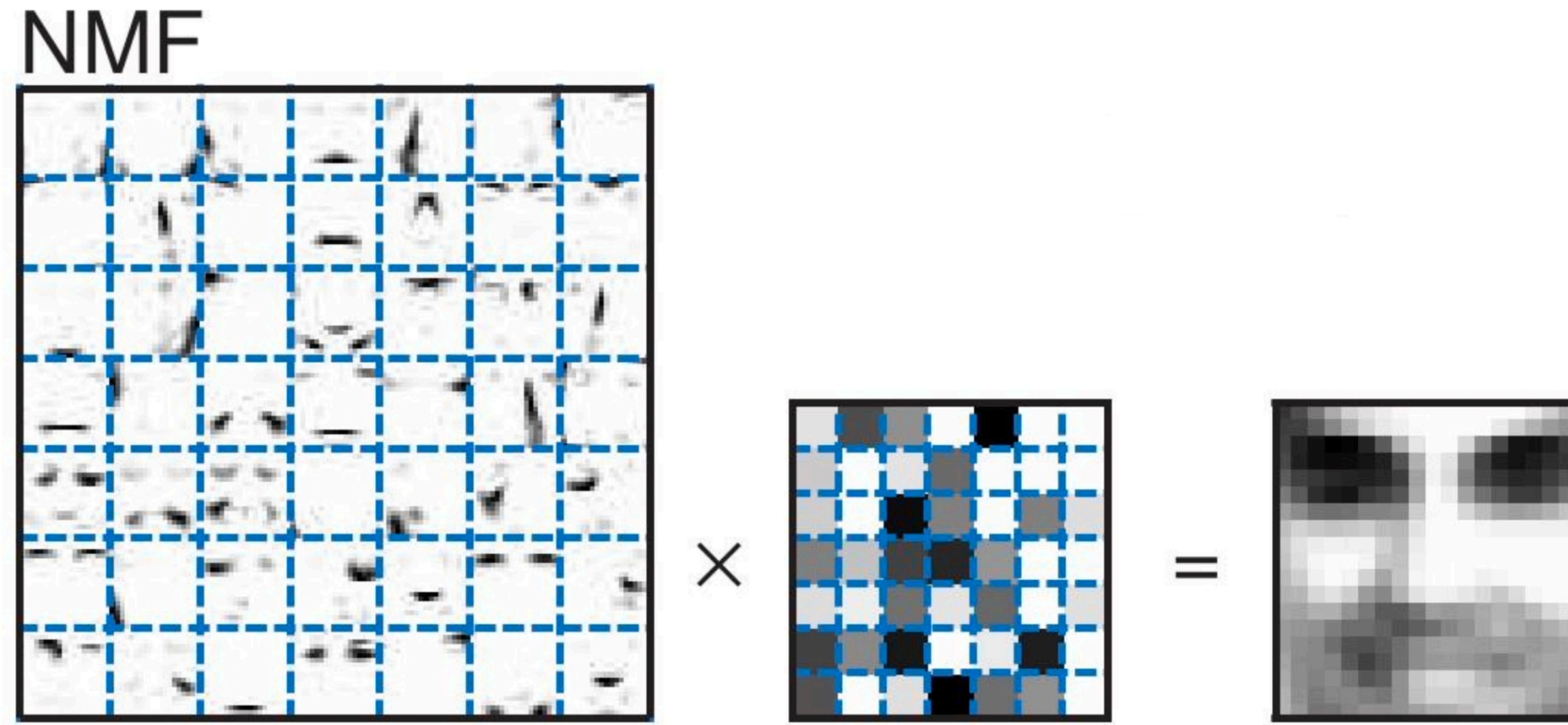
Feature Extraction & Dimensionality Reduction

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)



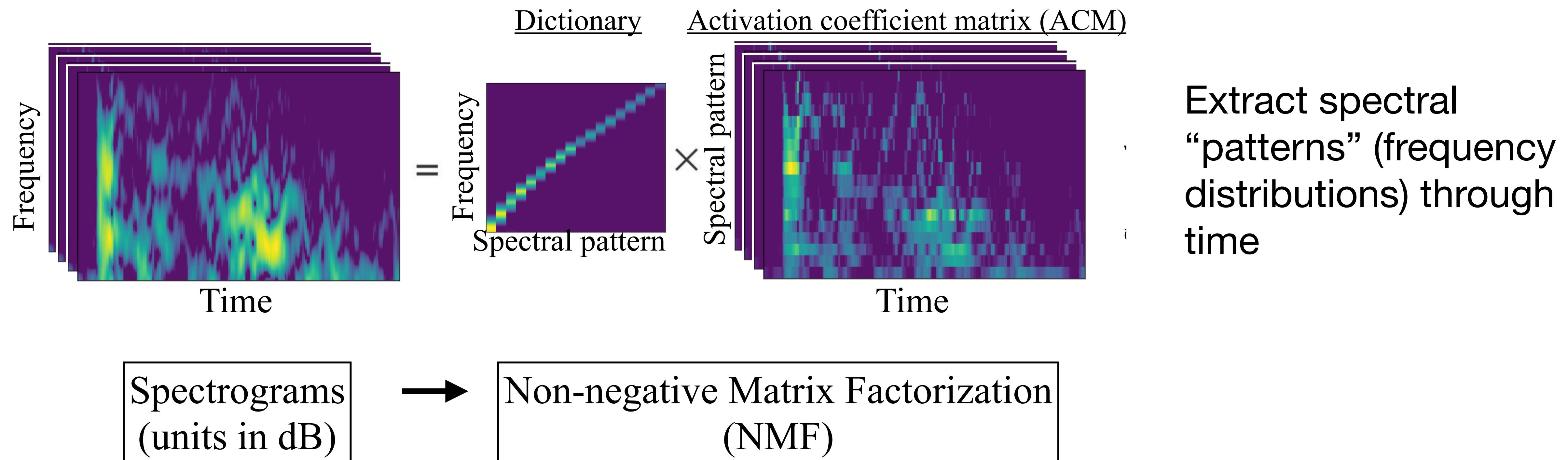
NMF learns a “parts-based” representation. Each column captures something interpretable. This is a result of the nonnegativity constraint.

Lee and H.S. Seung (2001)

Spectral Unsupervised Feature Extraction (SpecUFEx)

Holtzman et al., 2018; Sci Adv

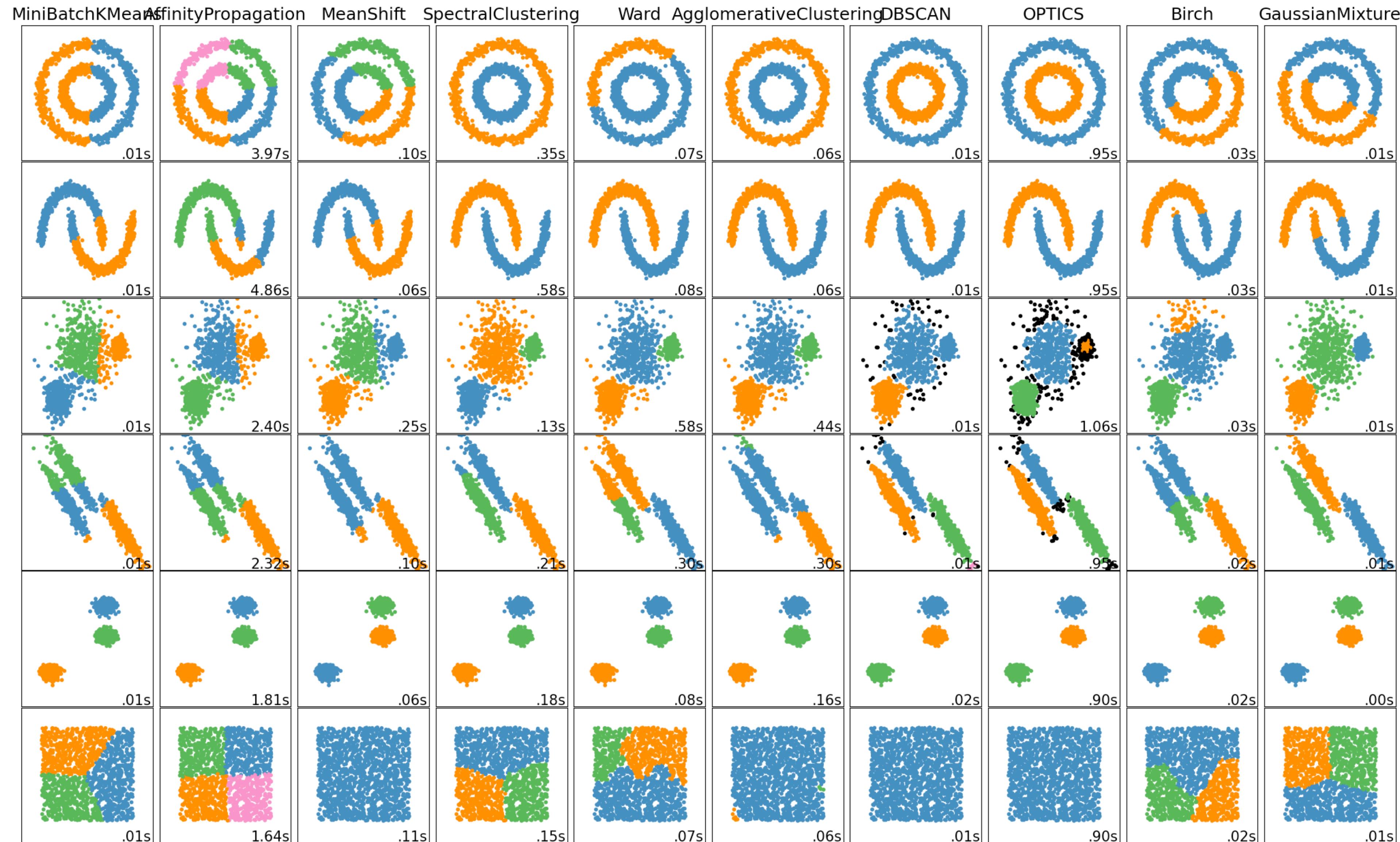
Step 1. Non-Negative Matrix Factorization



Outline

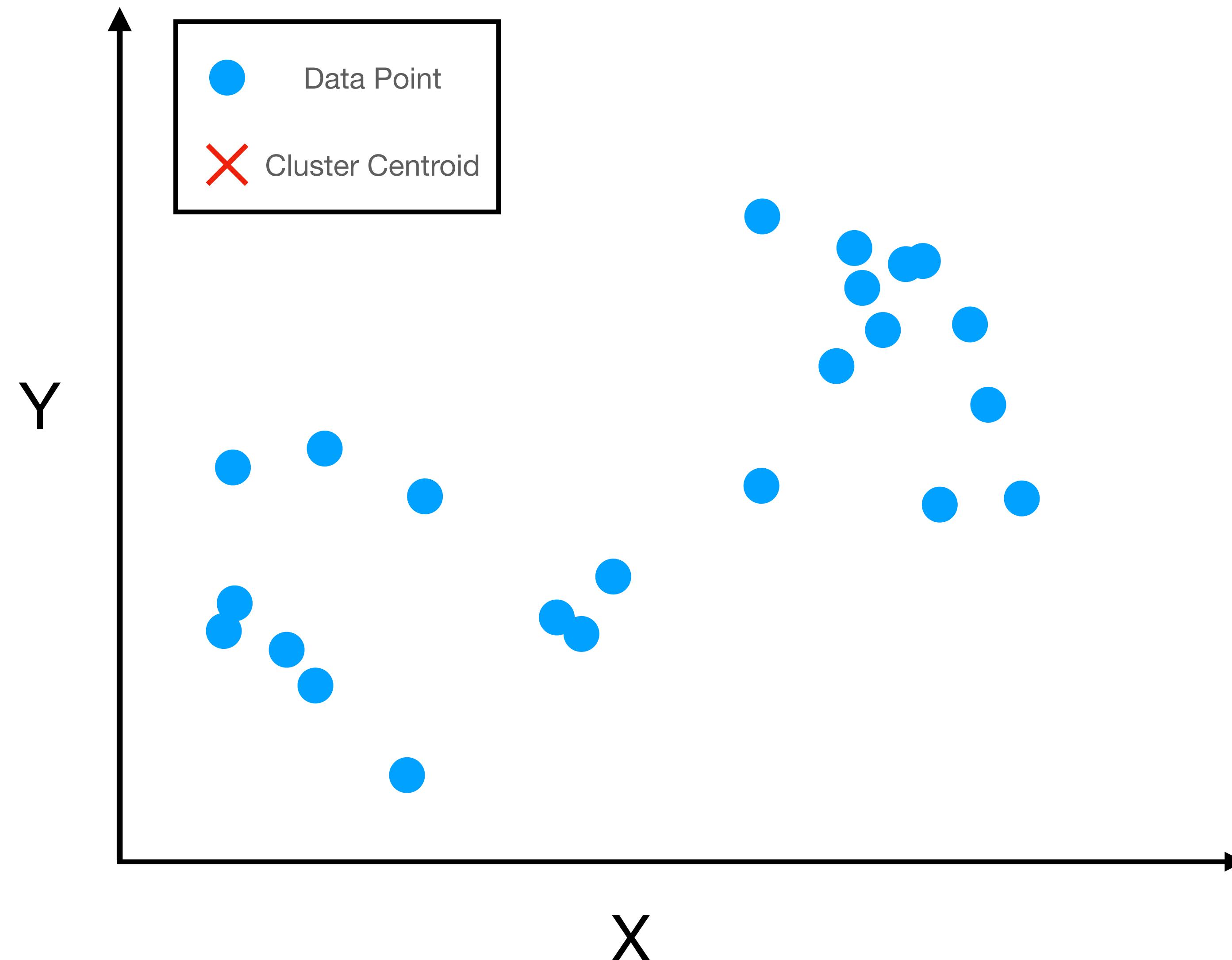
- Introduction to unsupervised machine learning (UML)
 - feature extraction & dimensionality reduction
 - **clustering**
- **SpecUFEx tutorial: Amatrice 2016**

Clustering



Clustering: K-means

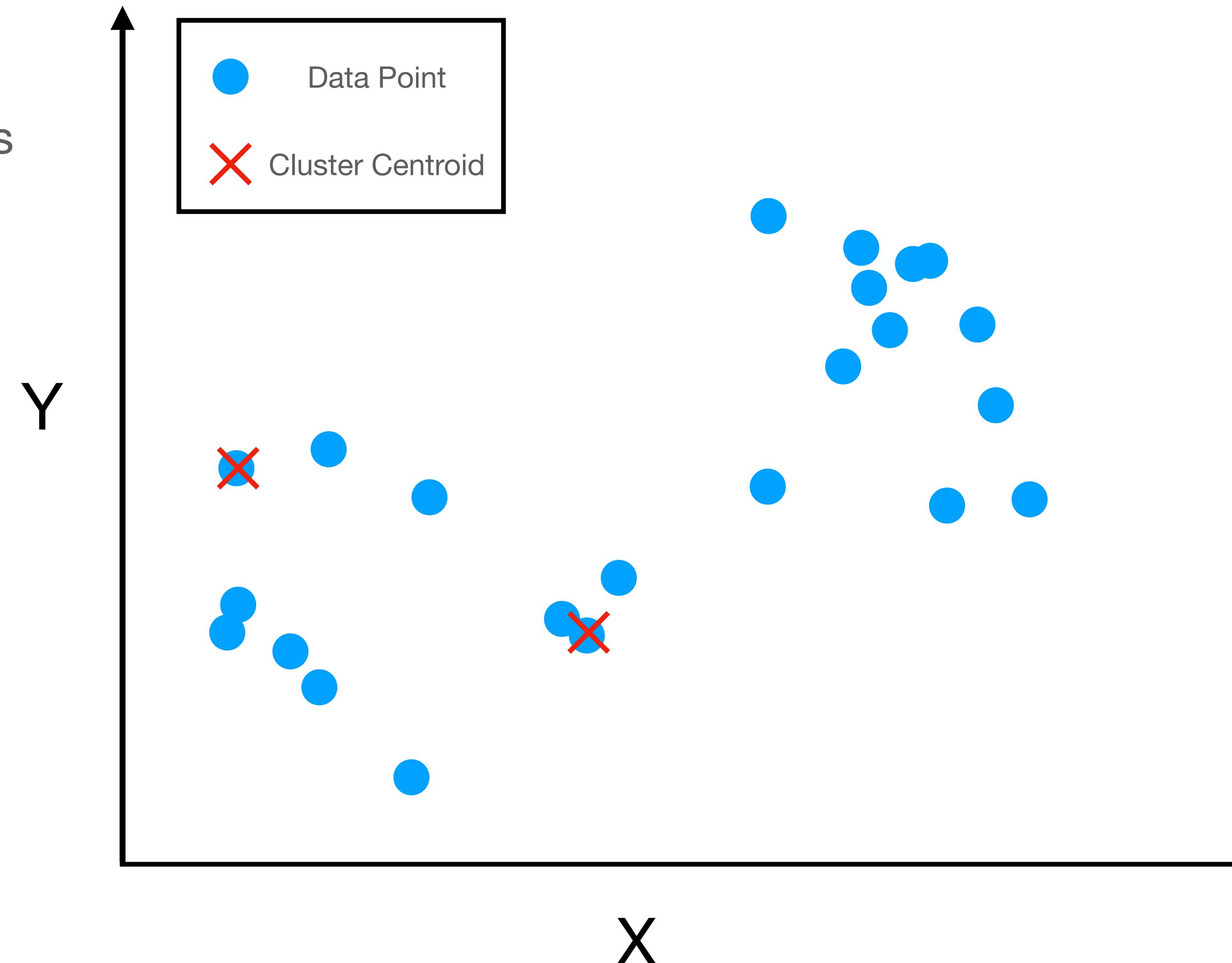
Lloyd, 1982, IEEE



Clustering: K-means

Lloyd, 1982, IEEE

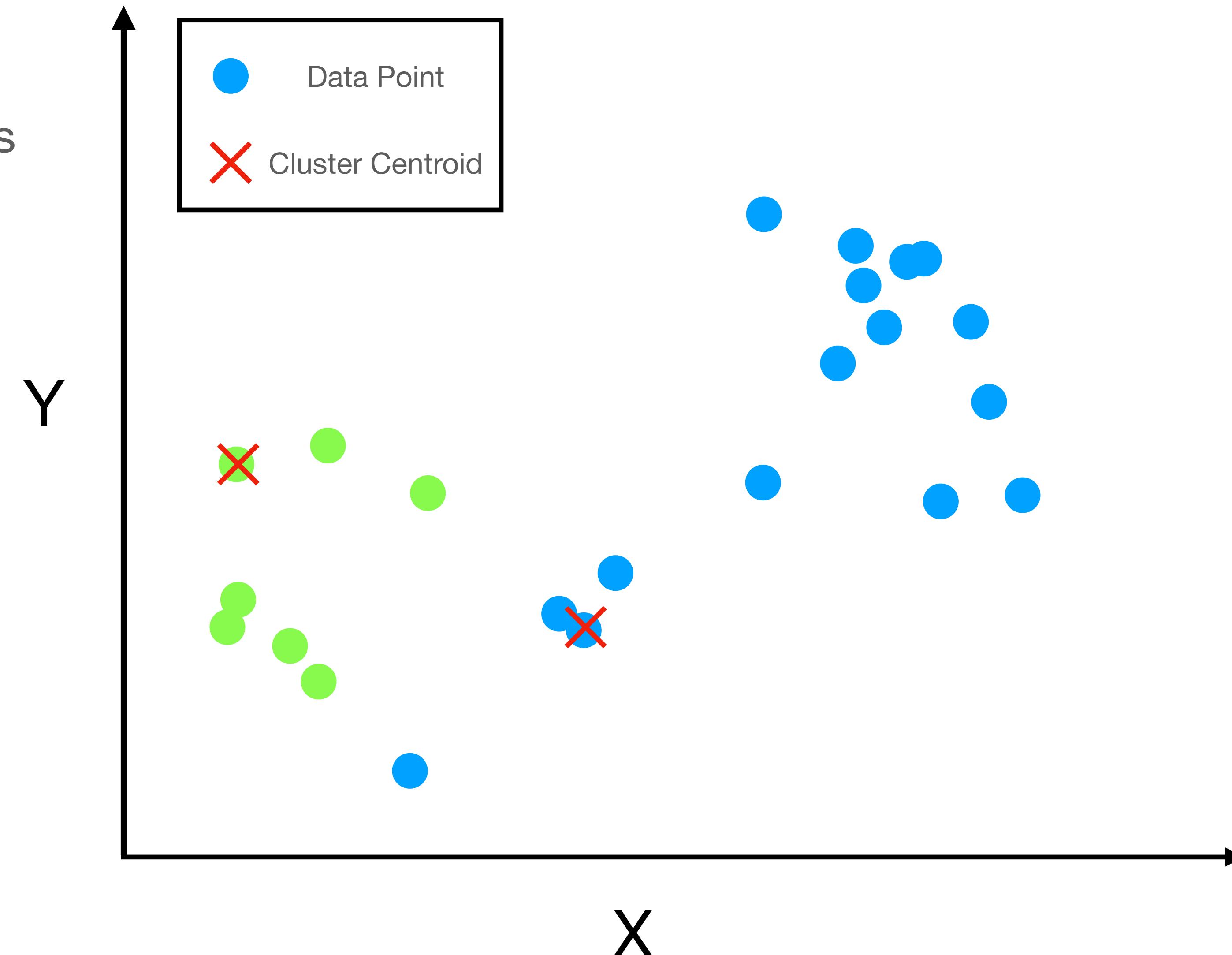
Step 1: Define centroids



Clustering: K-means

Lloyd, 1982, IEEE

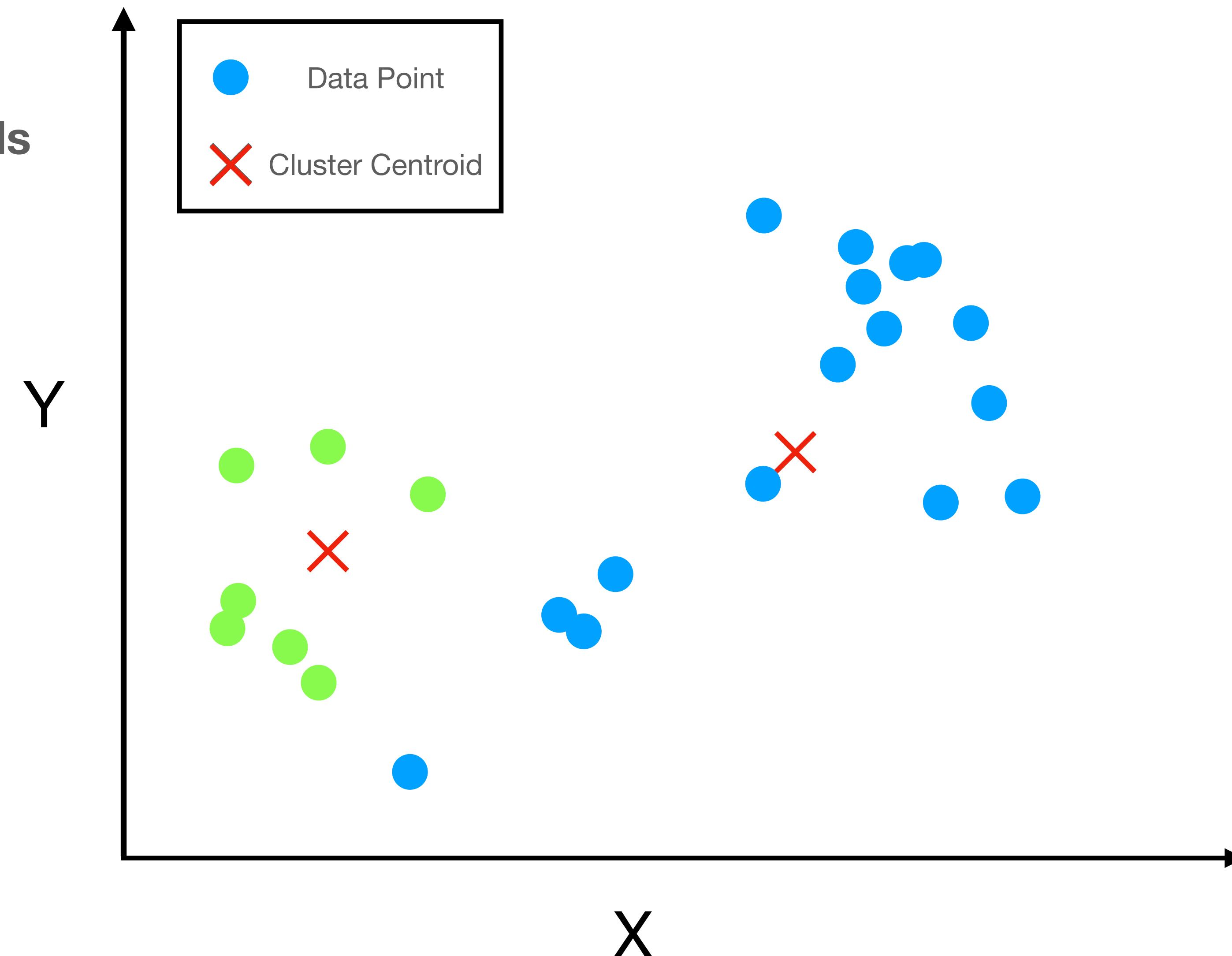
Step 1: Define centroids
Step 2: Cluster data



Clustering: K-means

Lloyd, 1982, IEEE

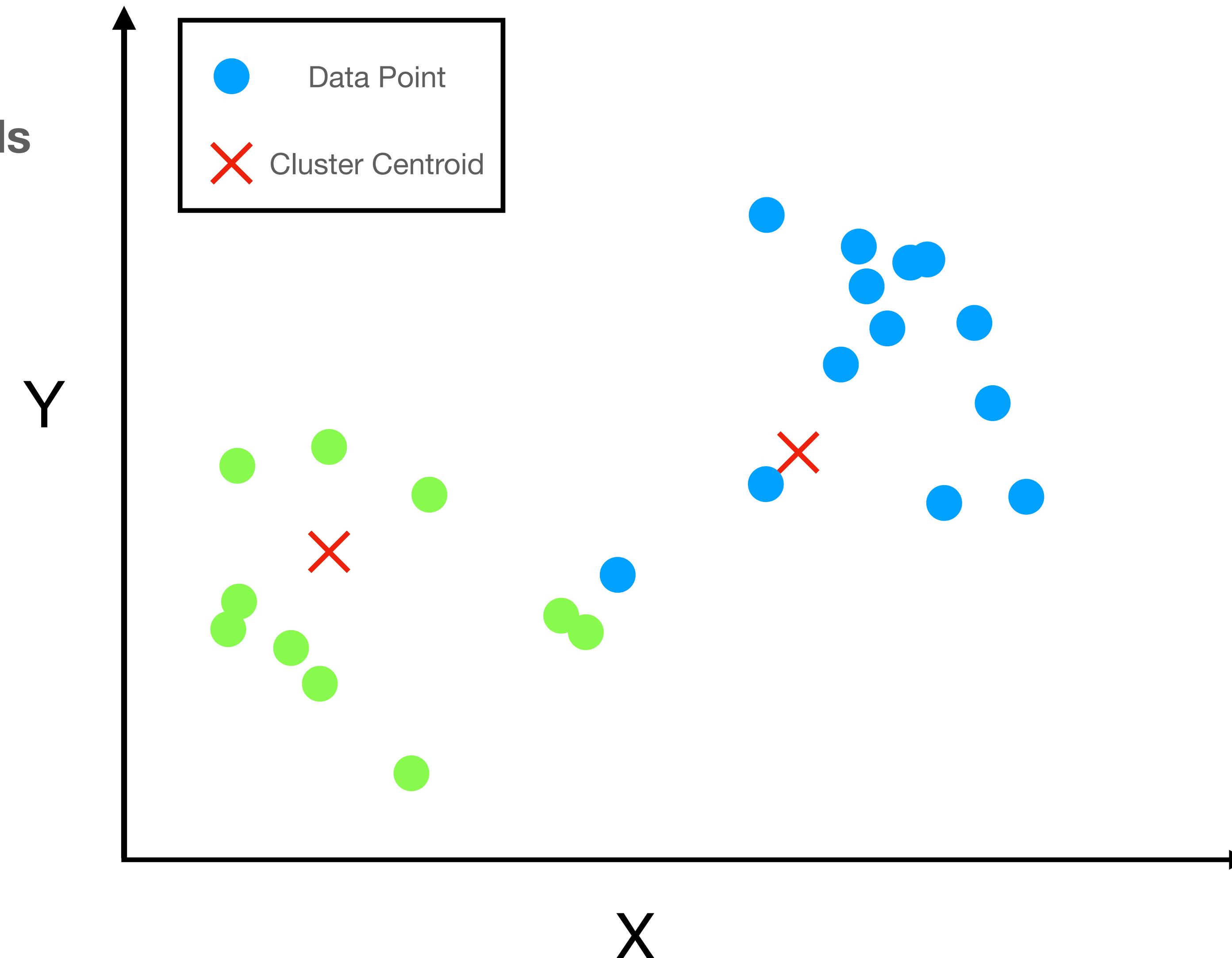
Step 1: Define centroids
Step 2: Cluster data



Clustering: K-means

Lloyd, 1982, IEEE

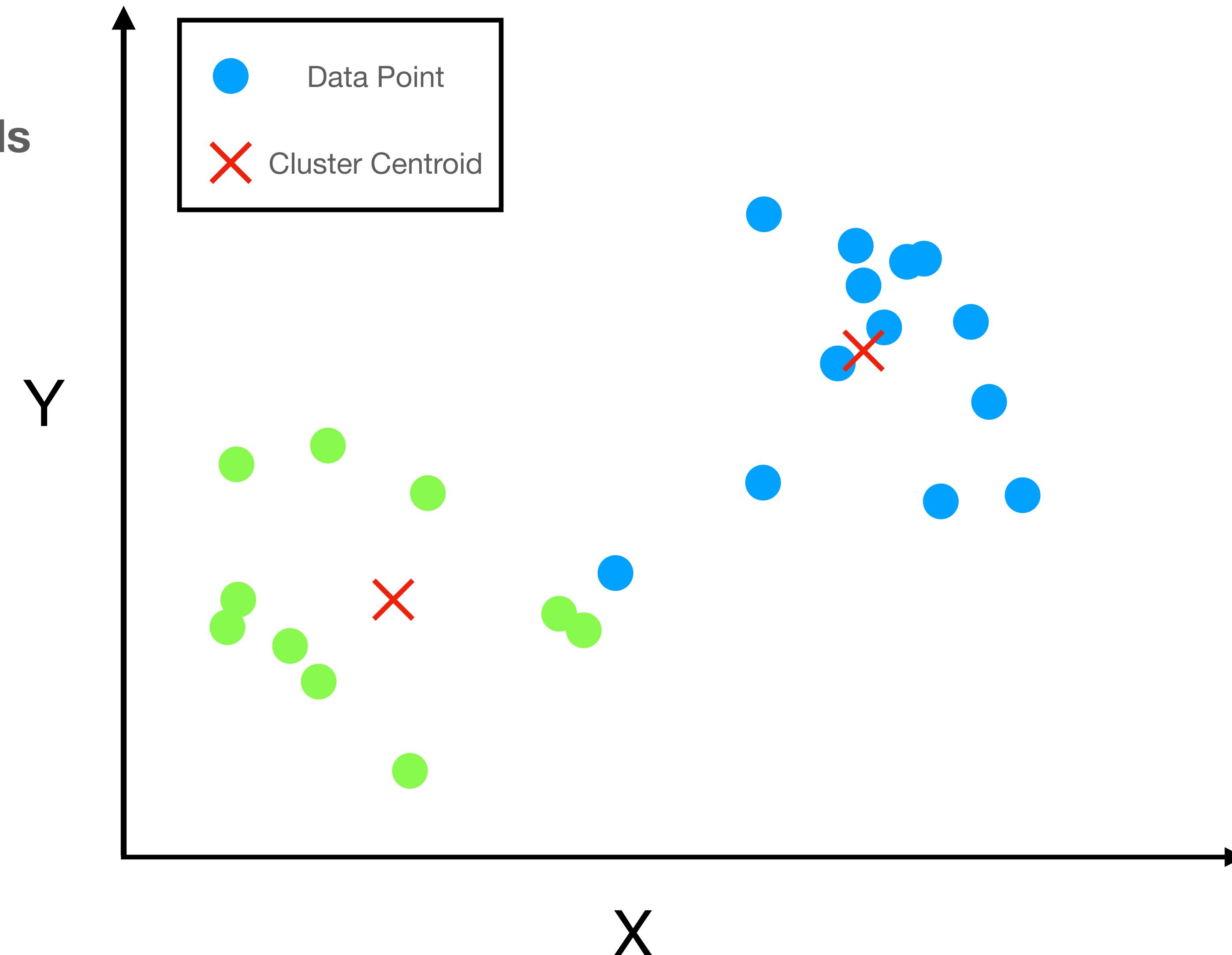
**Step 1: Define centroids
Step 2: Cluster data**



Clustering: K-means

Lloyd, 1982, IEEE

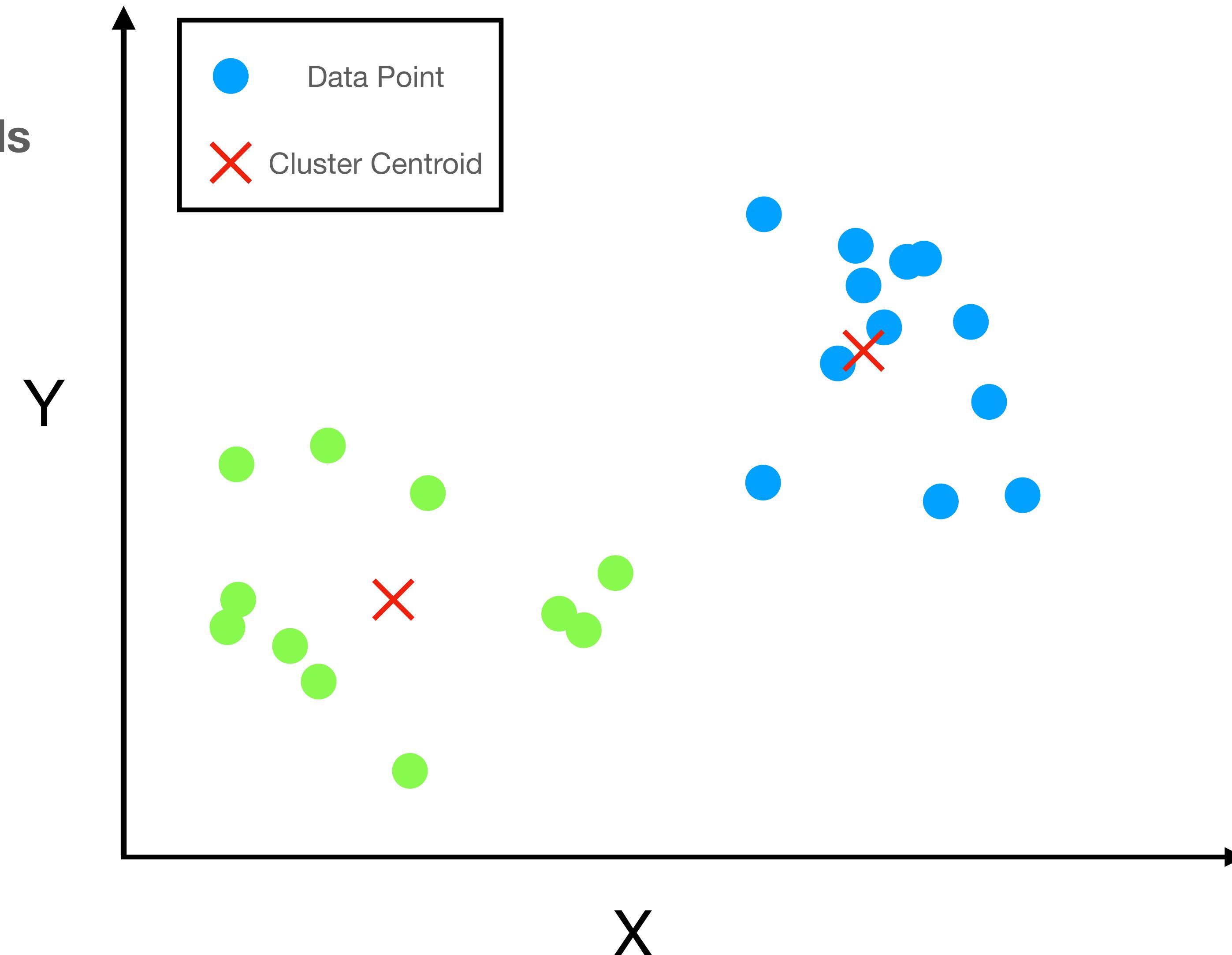
**Step 1: Define centroids
Step 2: Cluster data**



Clustering: K-means

Lloyd, 1982, IEEE

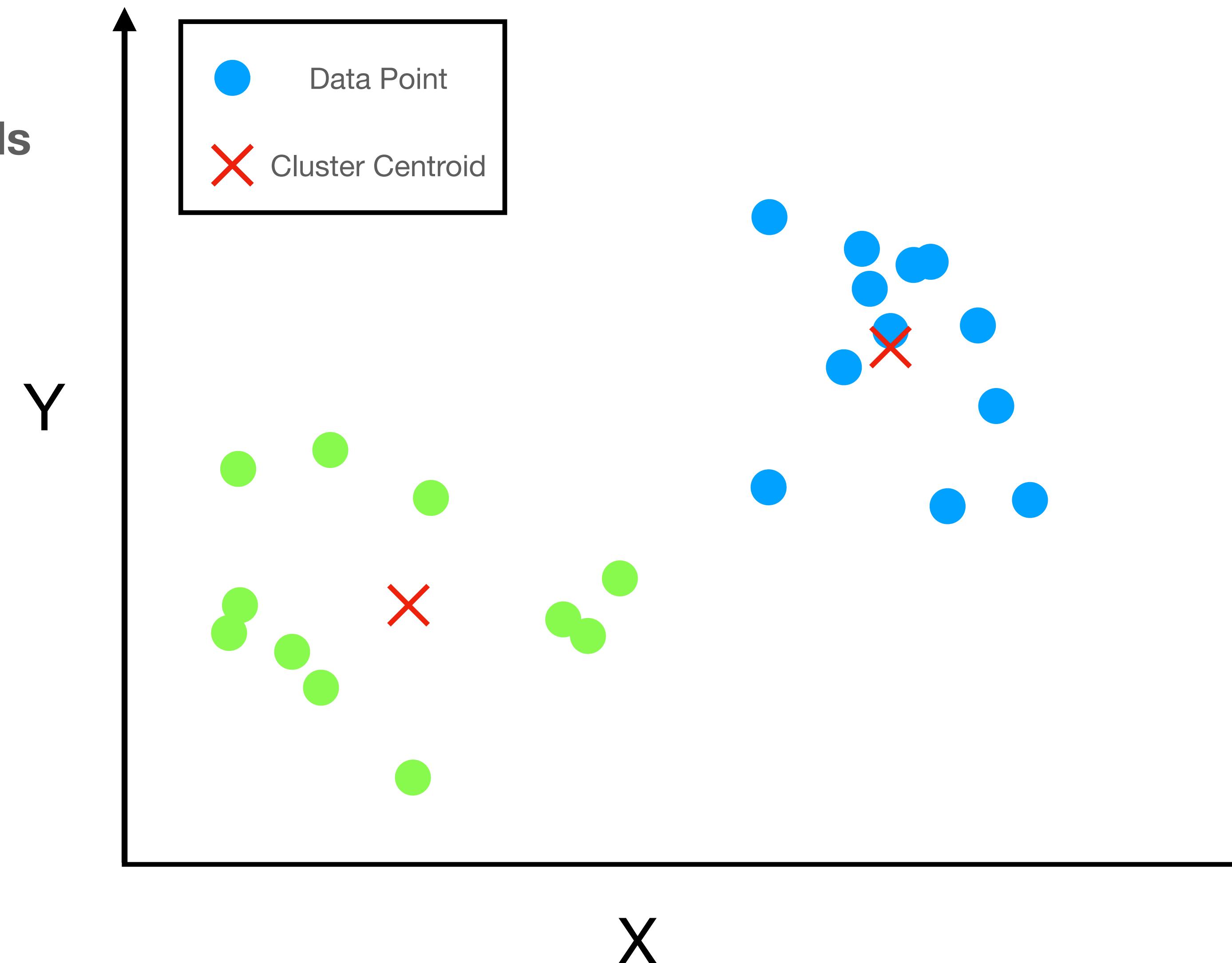
**Step 1: Define centroids
Step 2: Cluster data**



Clustering: K-means

Lloyd, 1982, IEEE

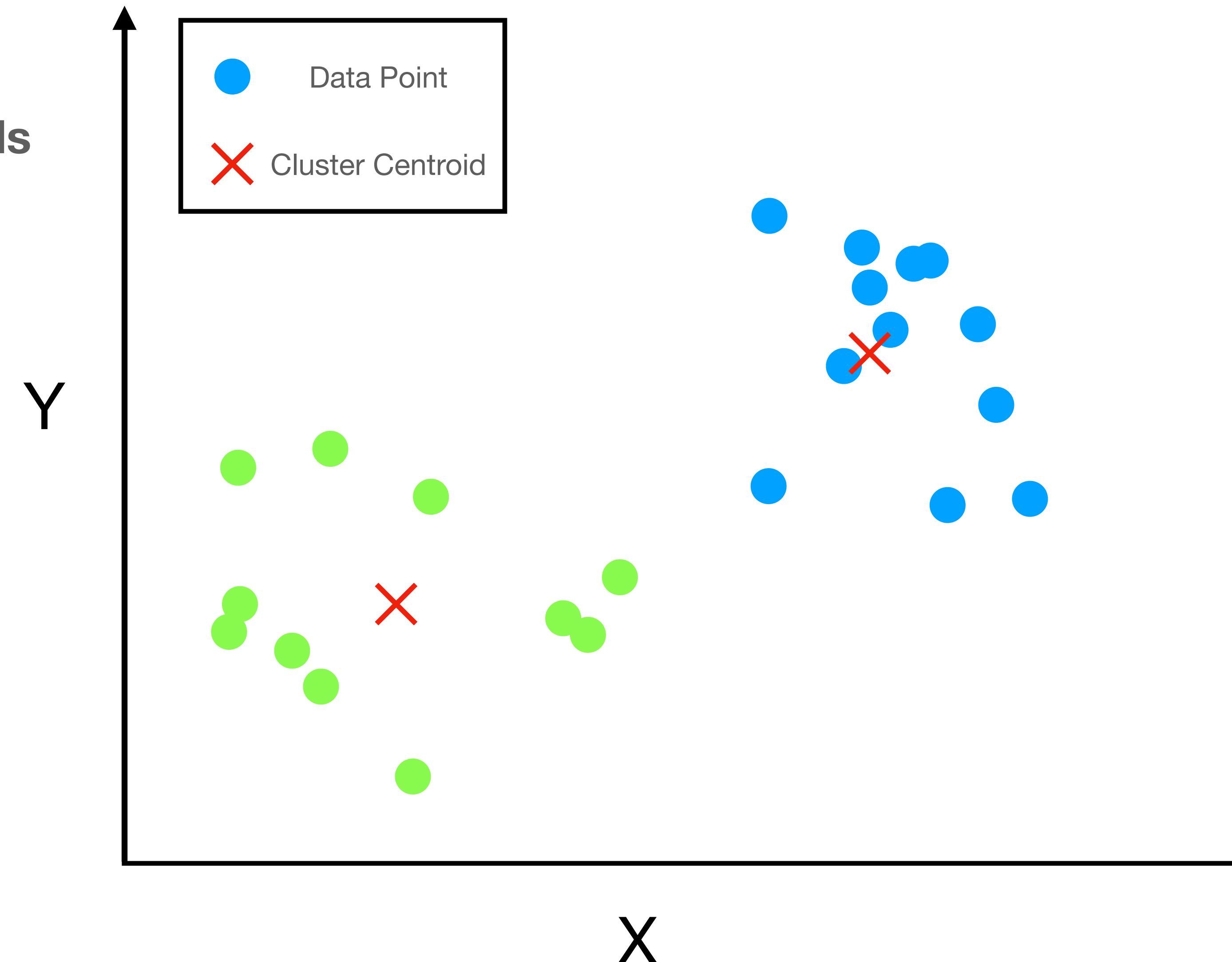
**Step 1: Define centroids
Step 2: Cluster data**



Clustering: K-means

Lloyd, 1982, IEEE

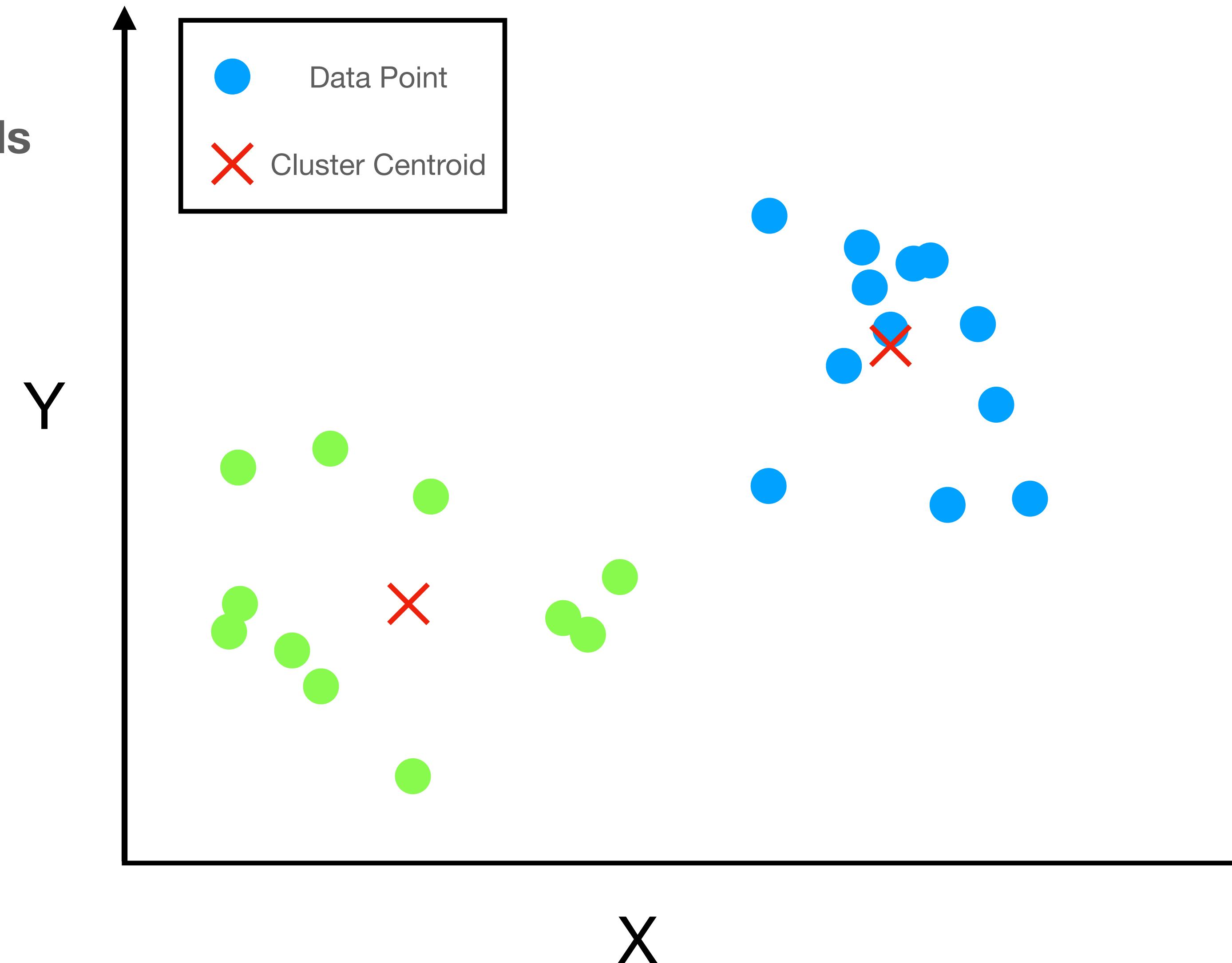
**Step 1: Define centroids
Step 2: Cluster data**



Clustering: K-means

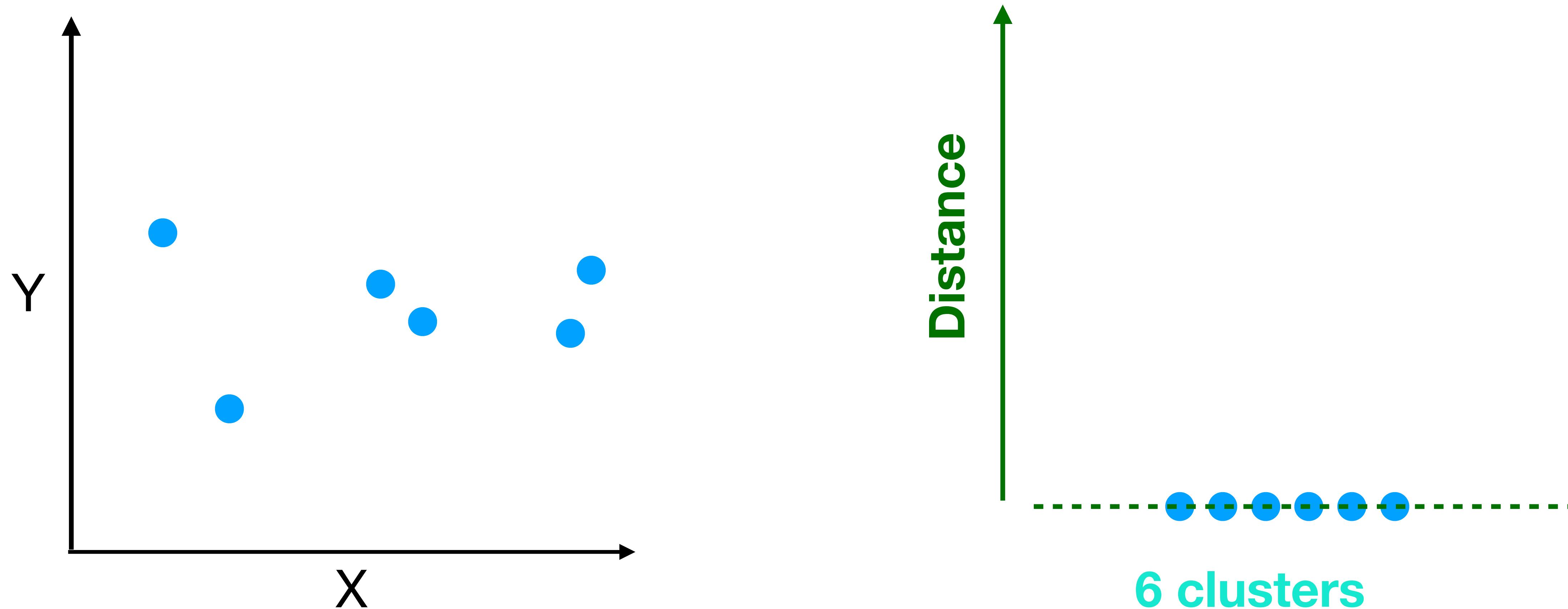
Lloyd, 1982, IEEE

**Step 1: Define centroids
Step 2: Cluster data**



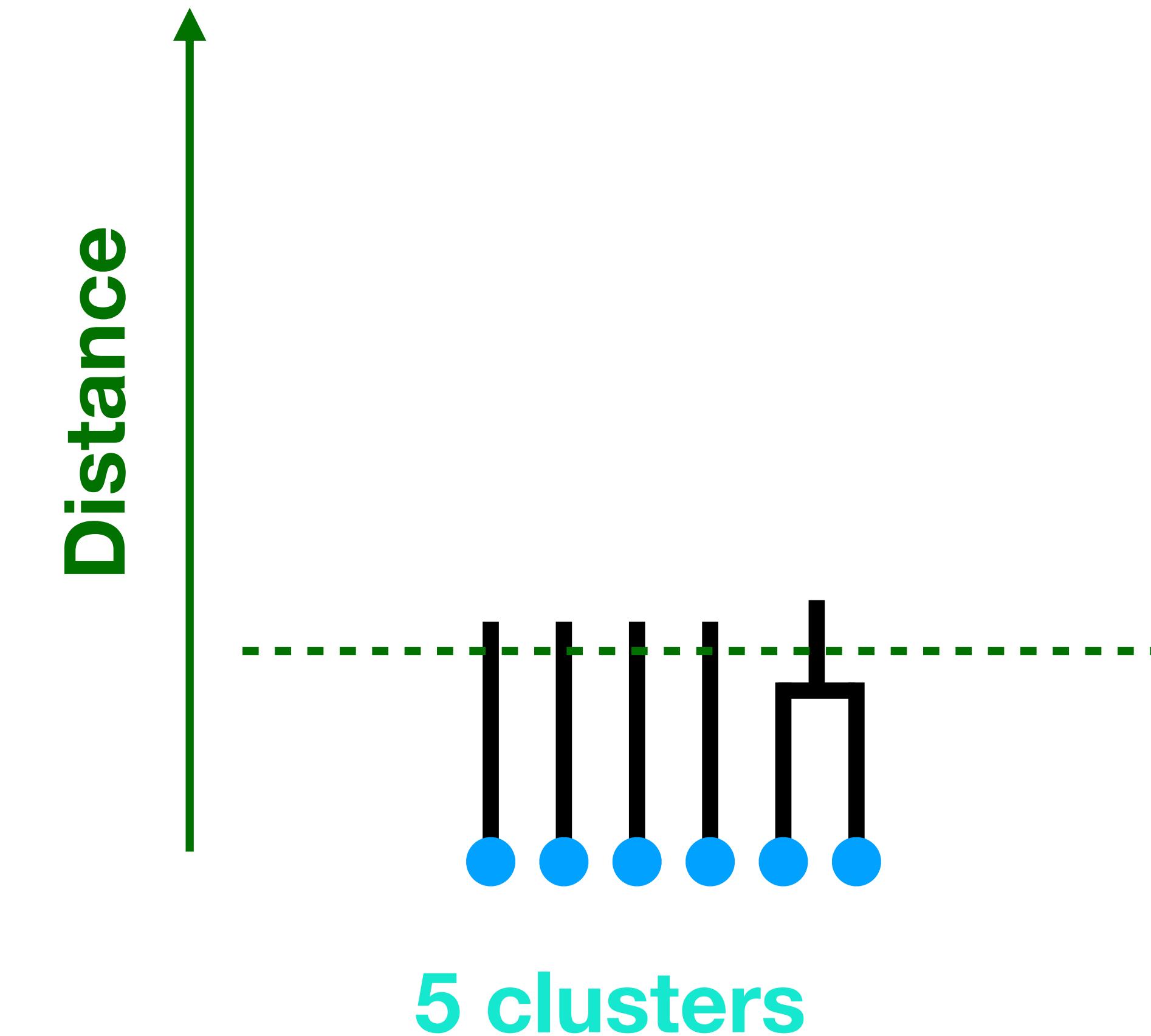
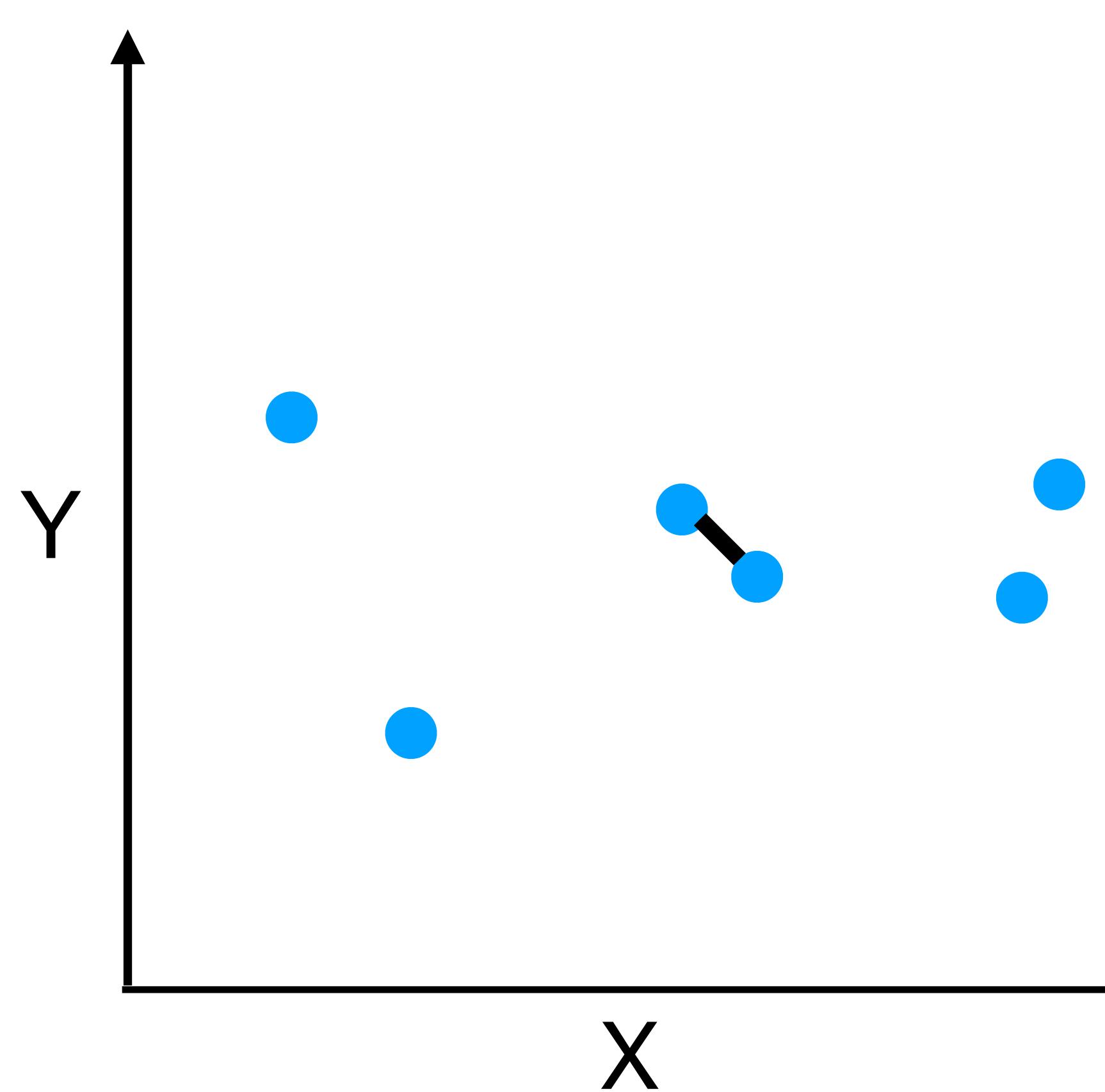
Hierarchical Clustering

w/ Single Linkage



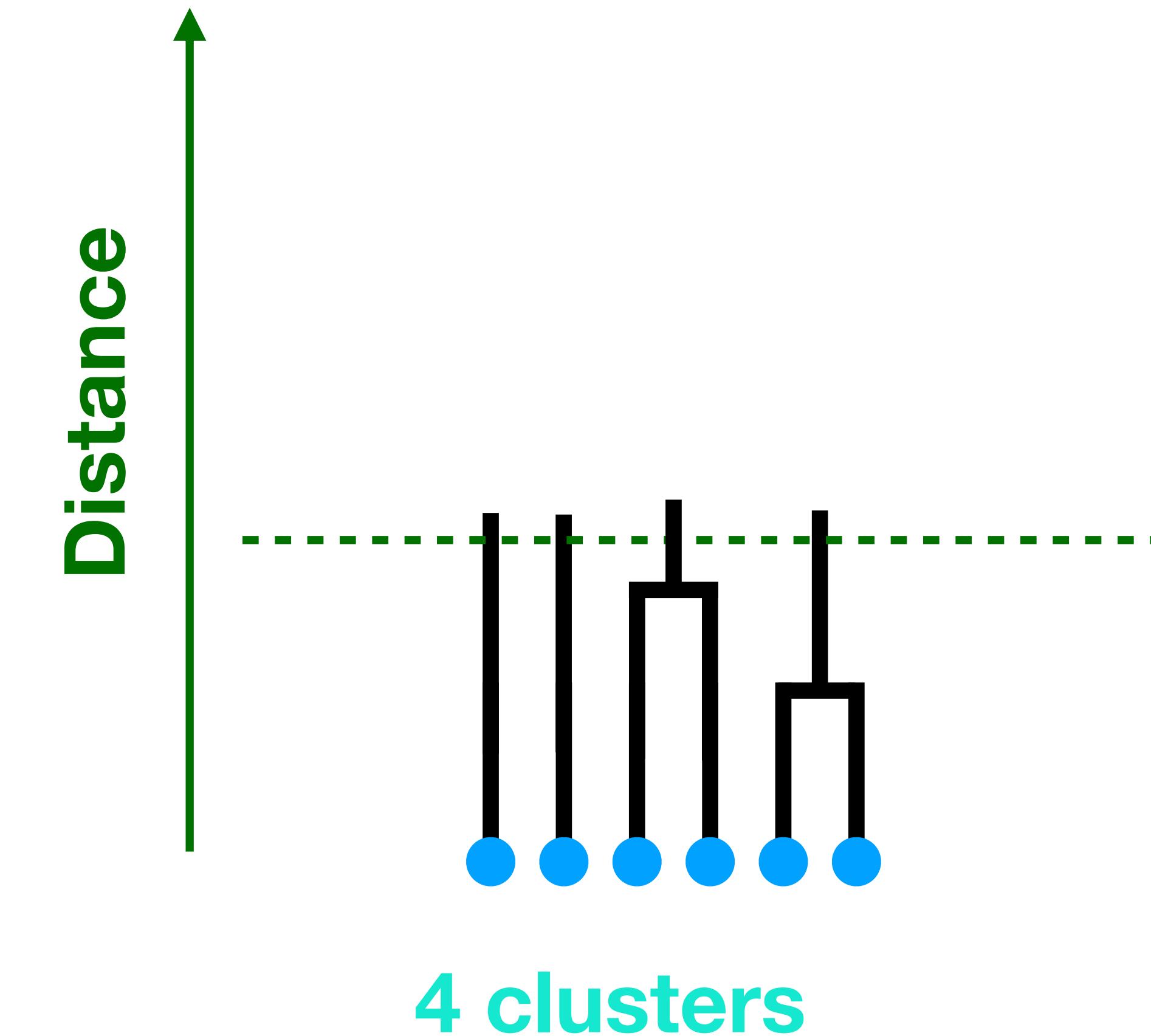
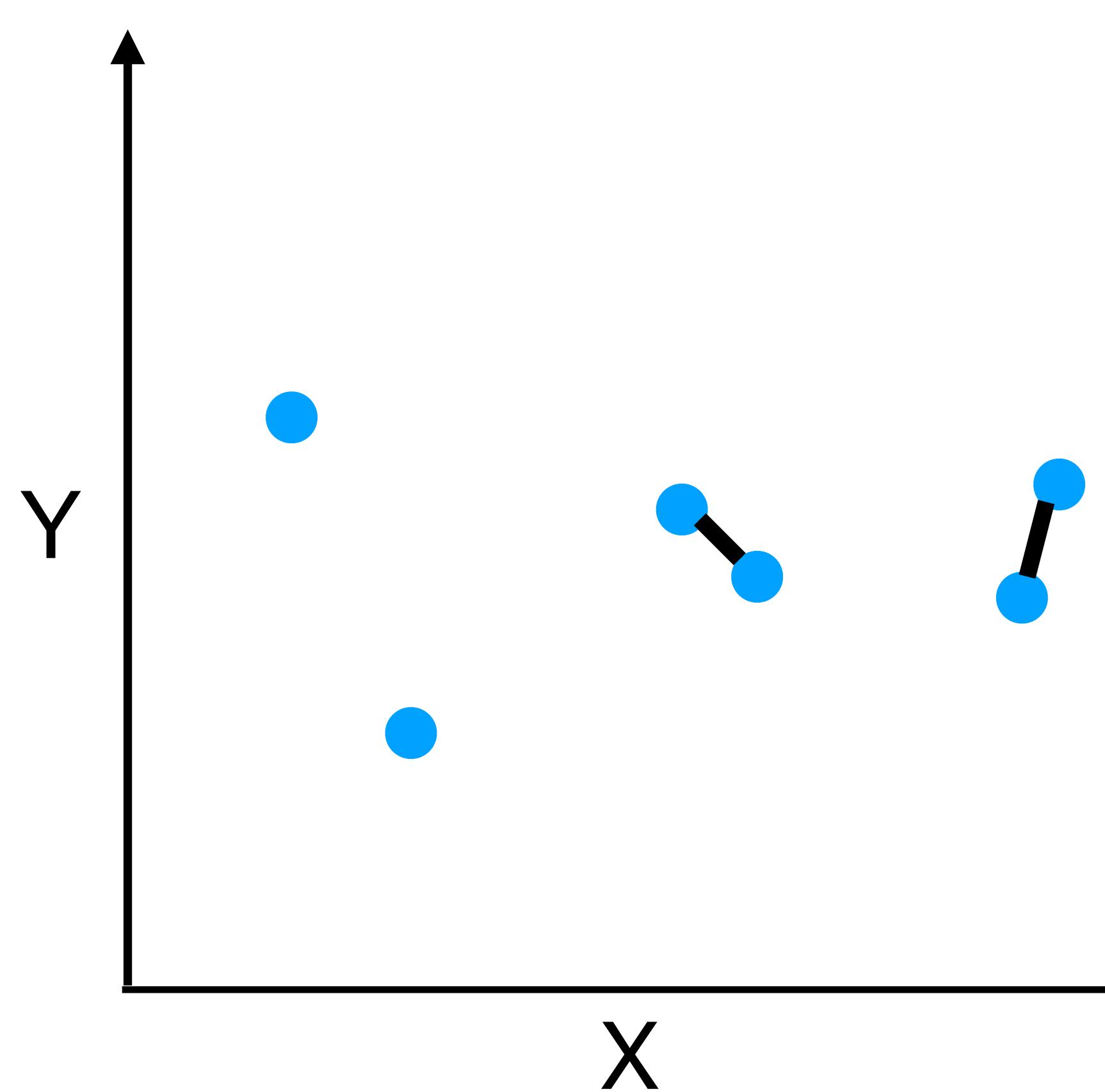
Hierarchical Clustering

w/ Single Linkage



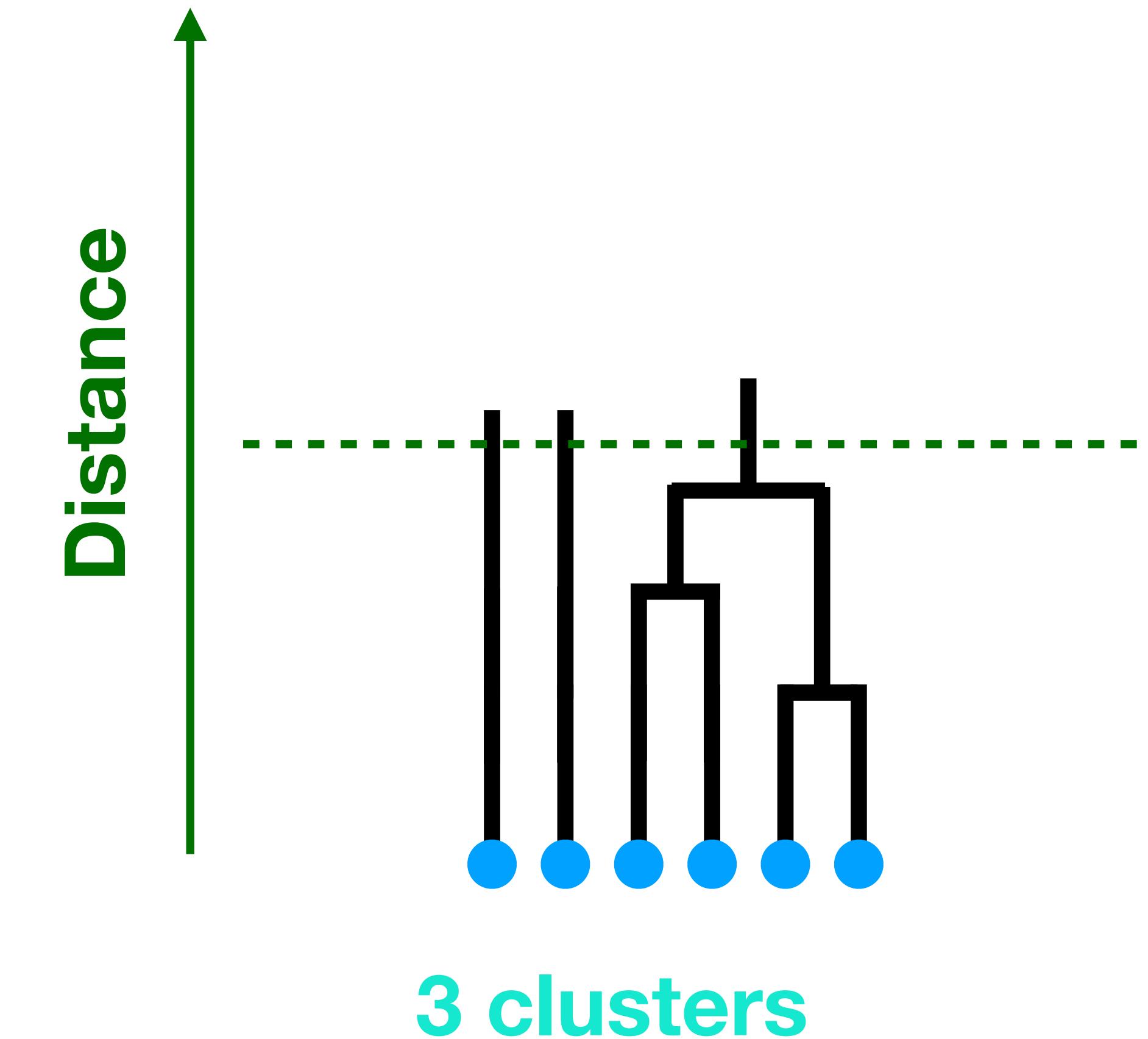
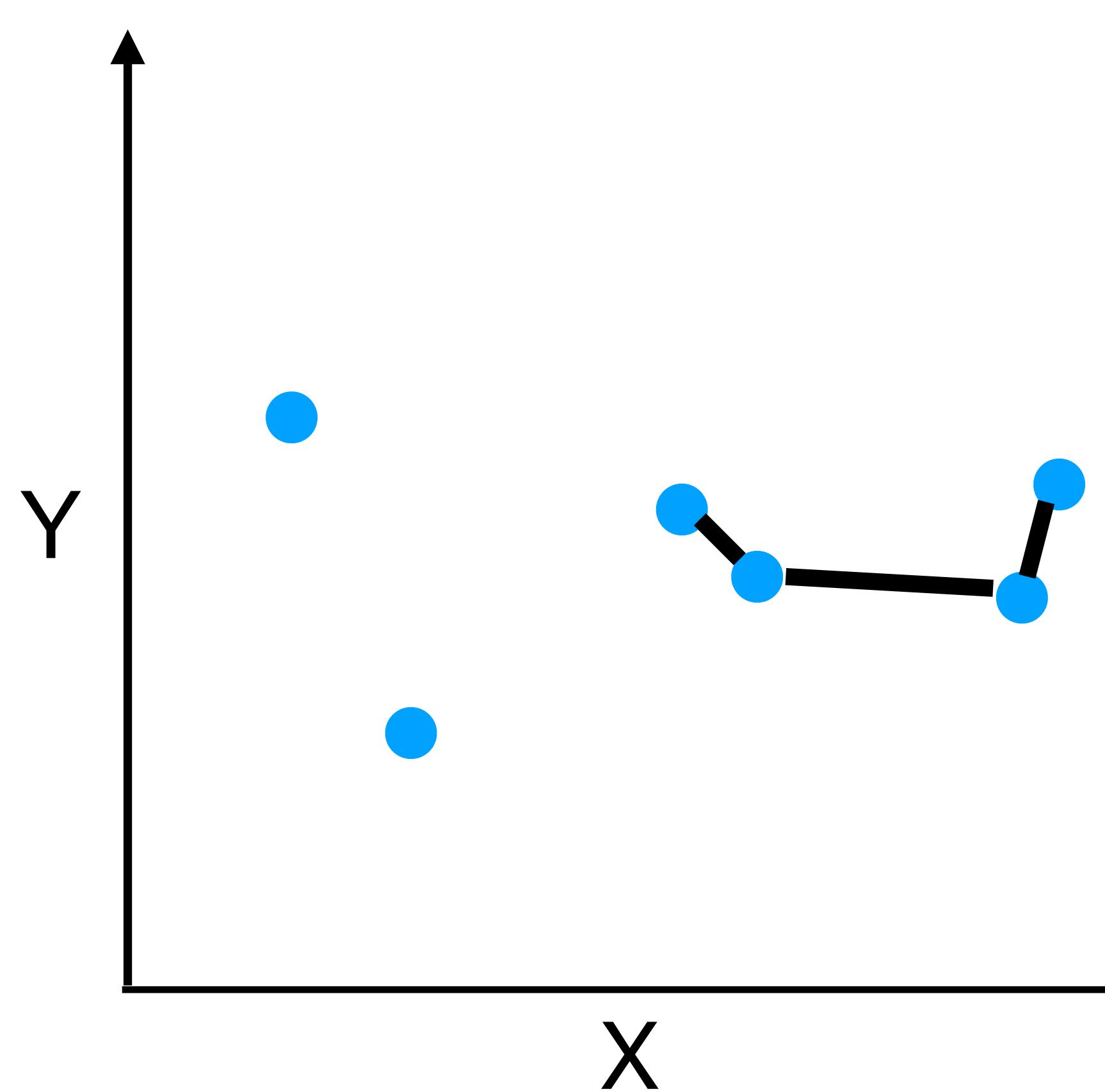
Hierarchical Clustering

w/ Single Linkage



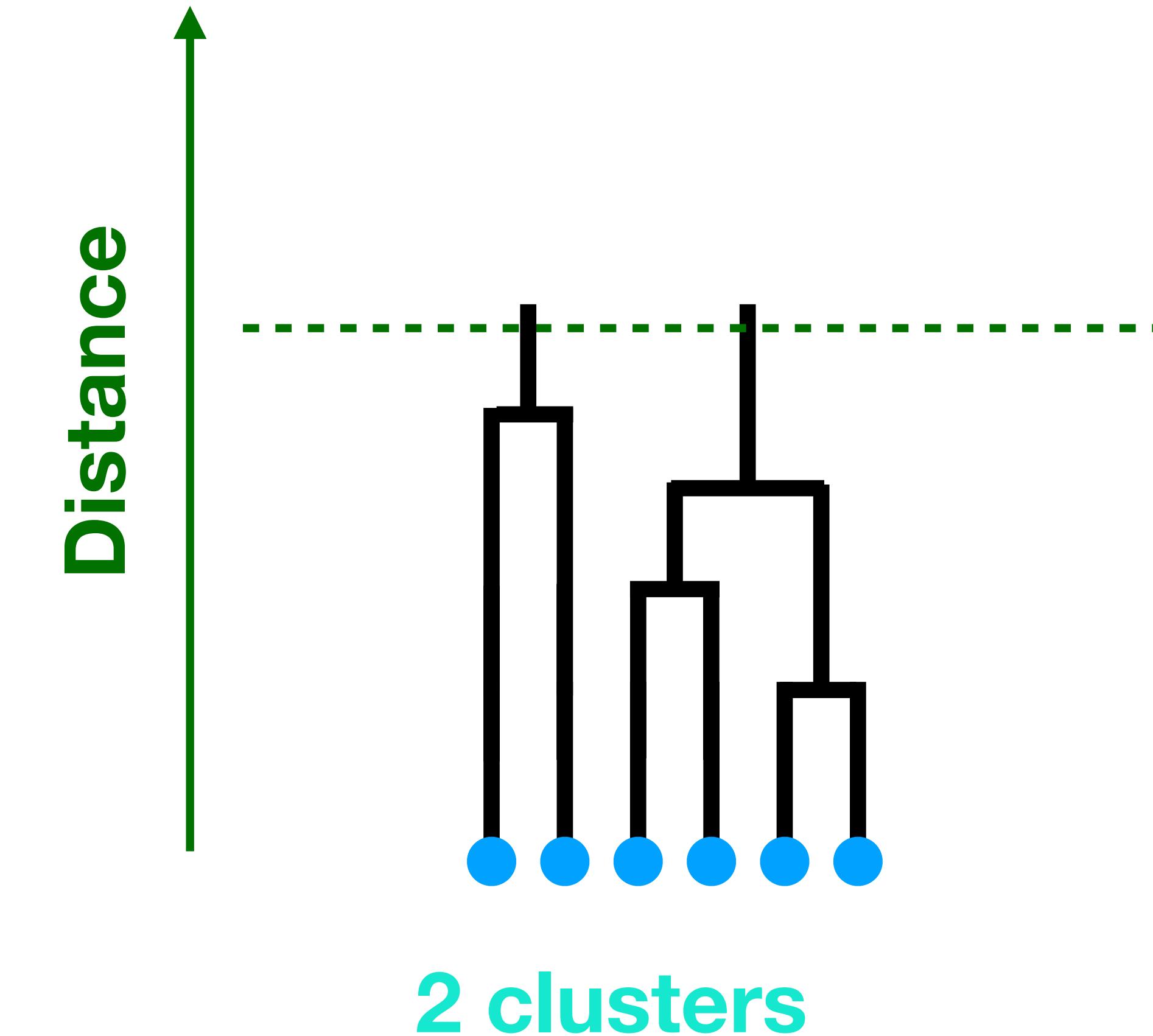
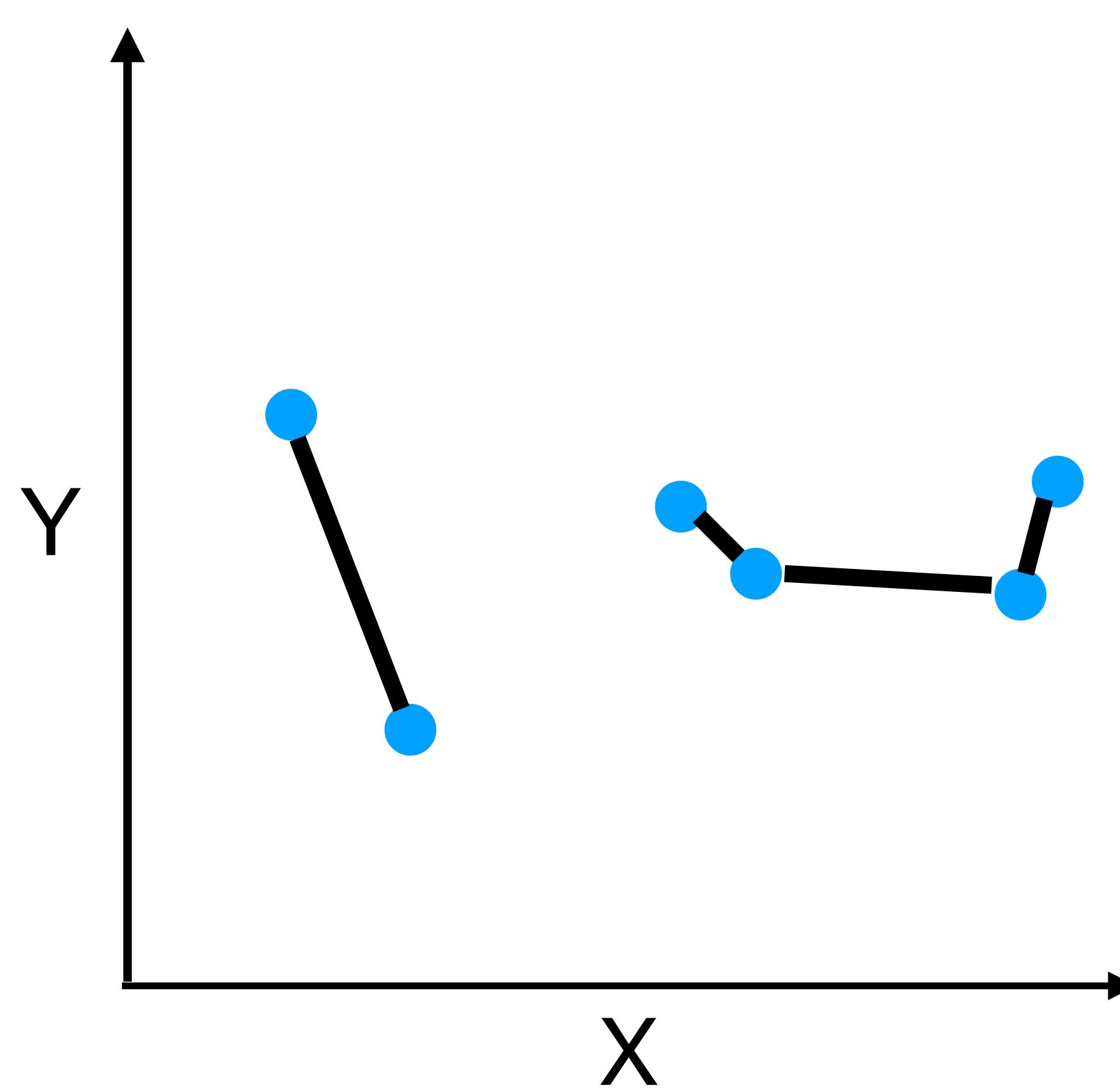
Hierarchical Clustering

w/ Single Linkage



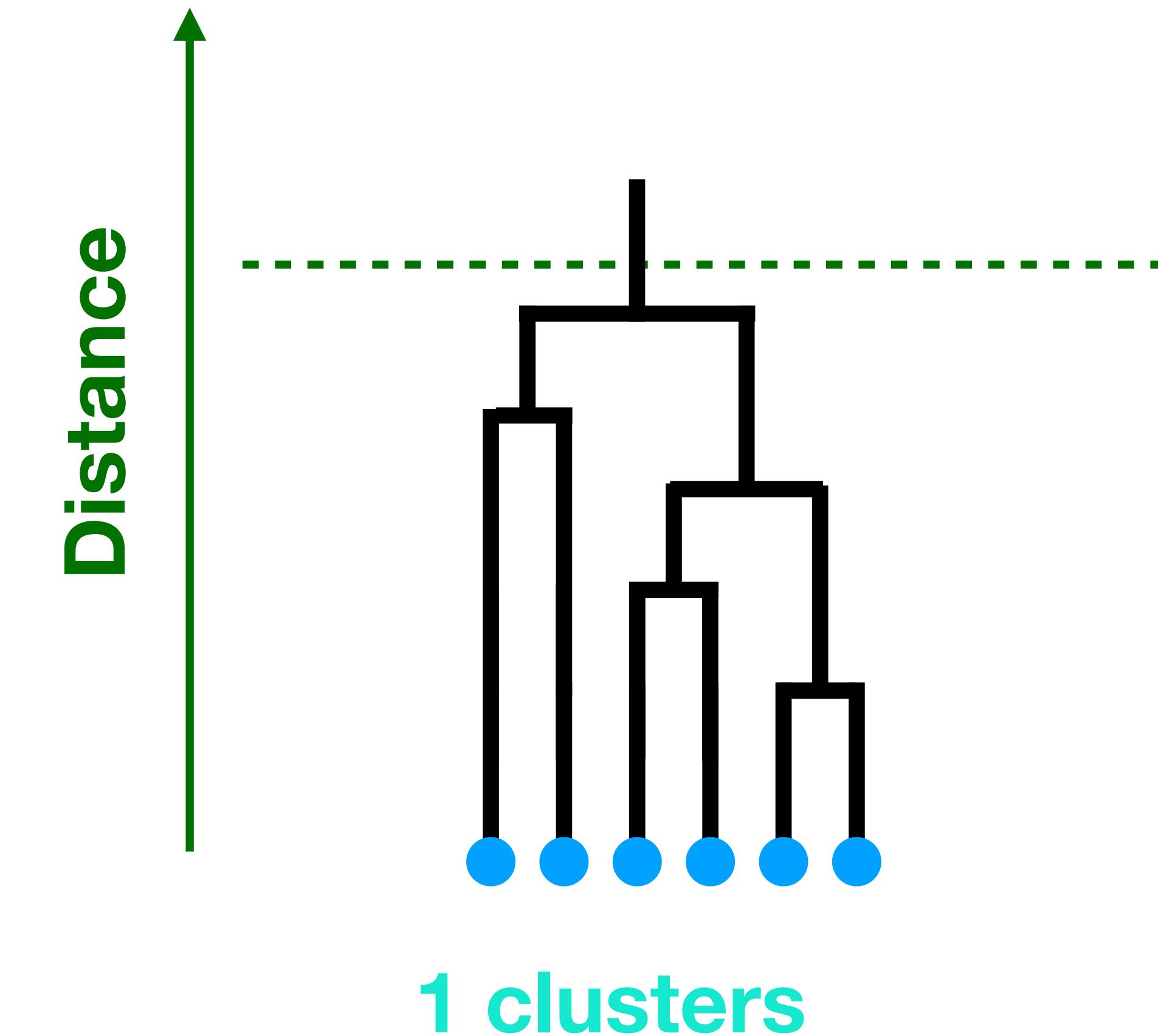
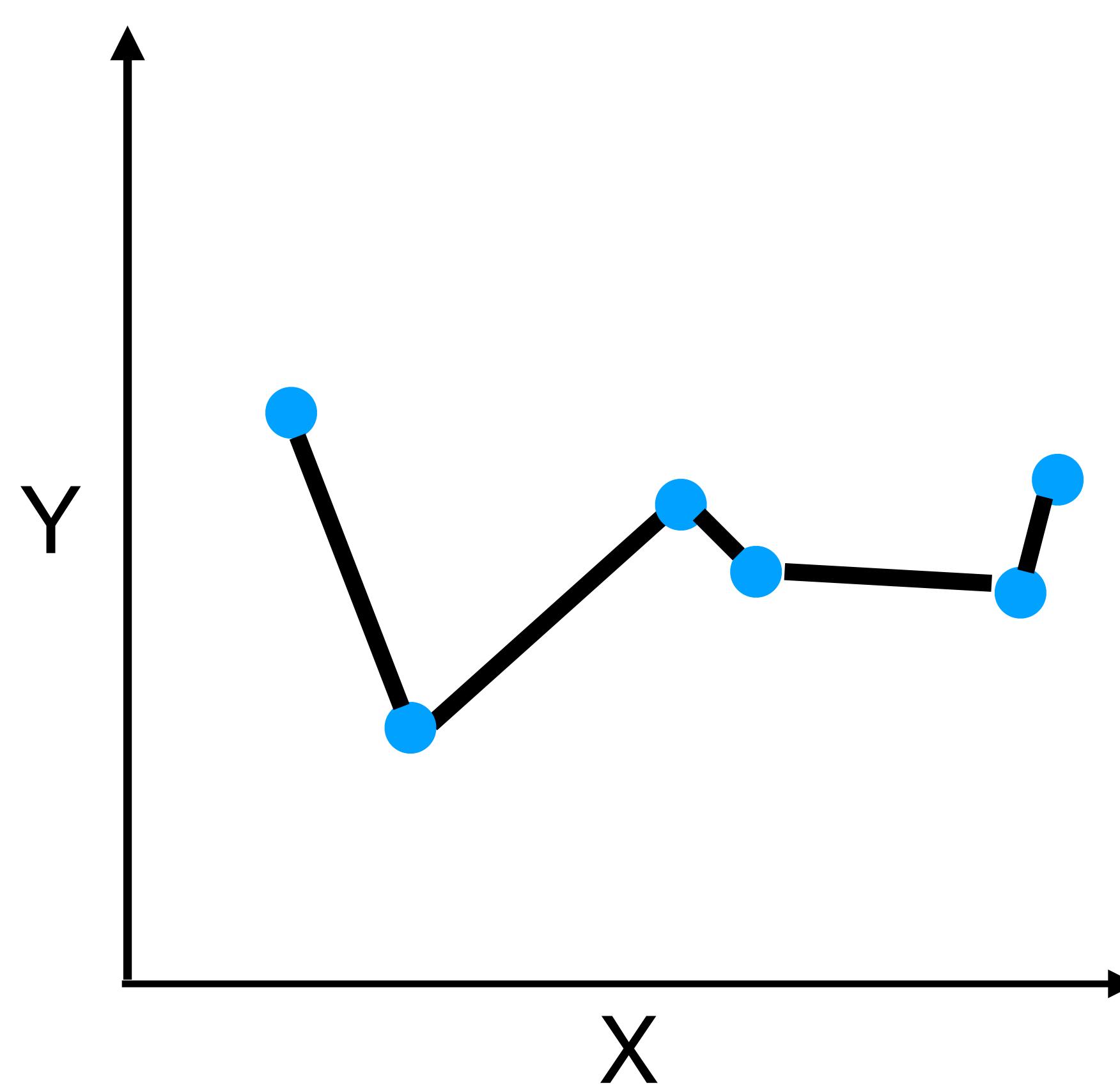
Hierarchical Clustering

w/ Single Linkage



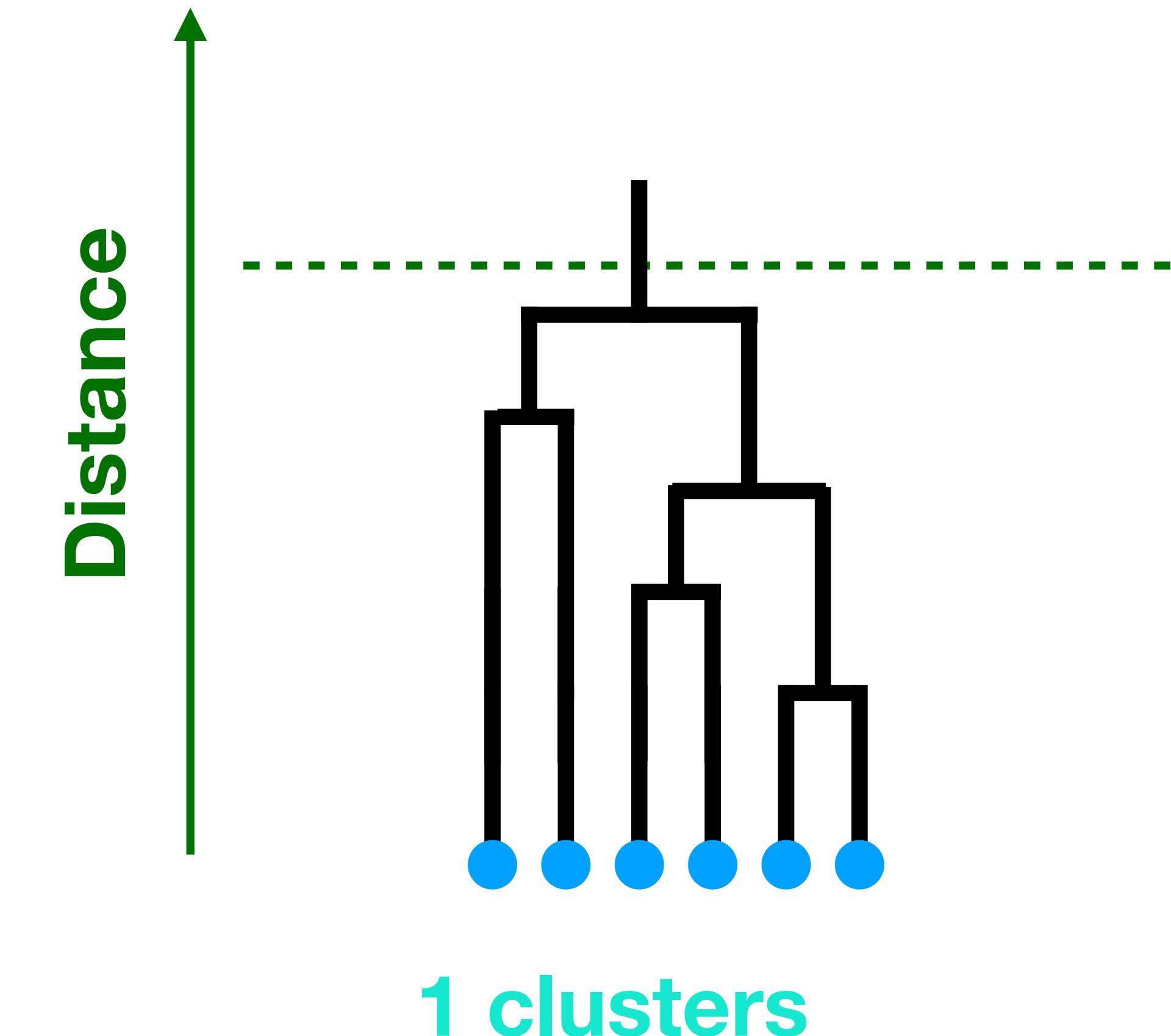
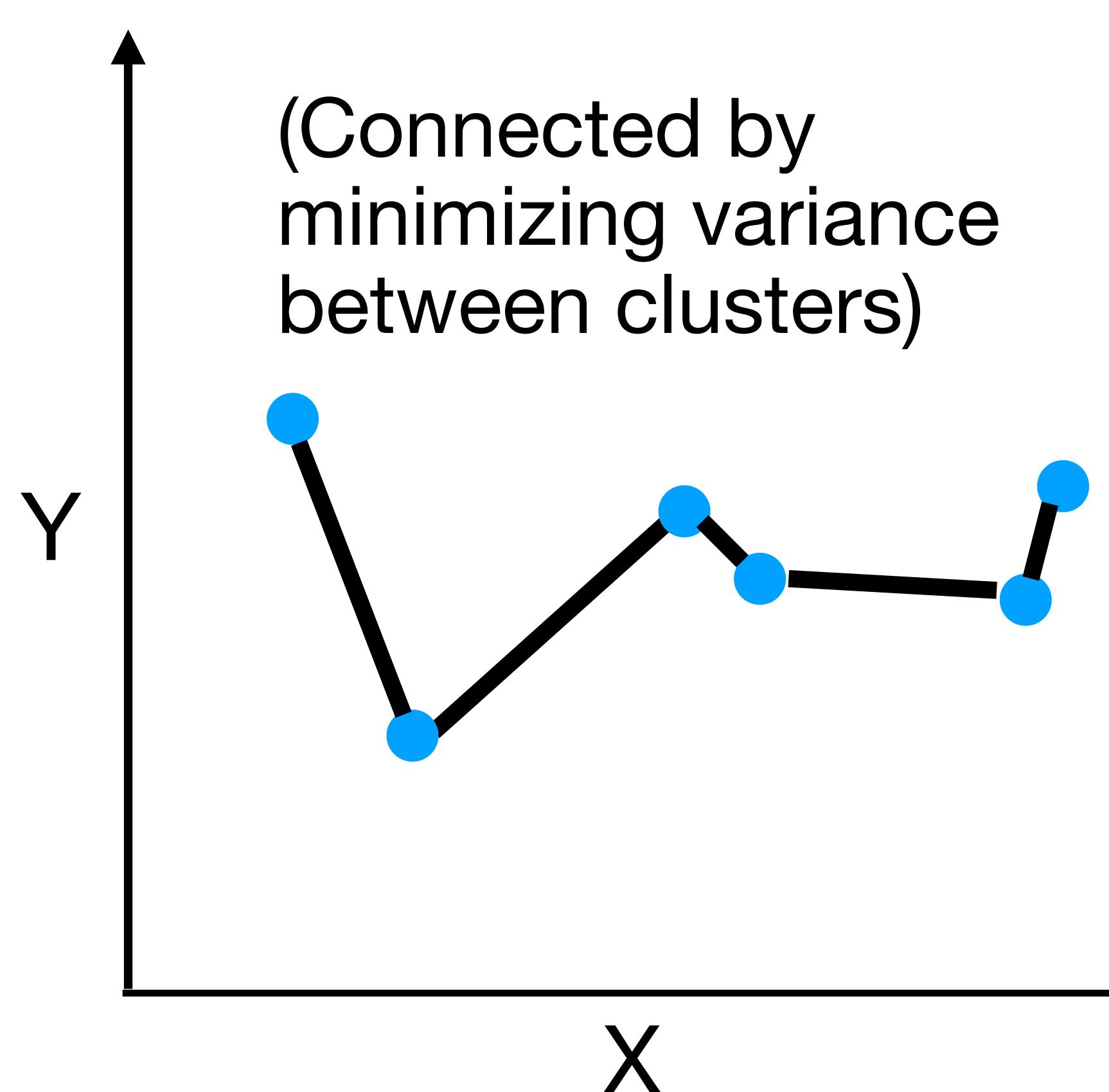
Hierarchical Clustering

w/ Single Linkage

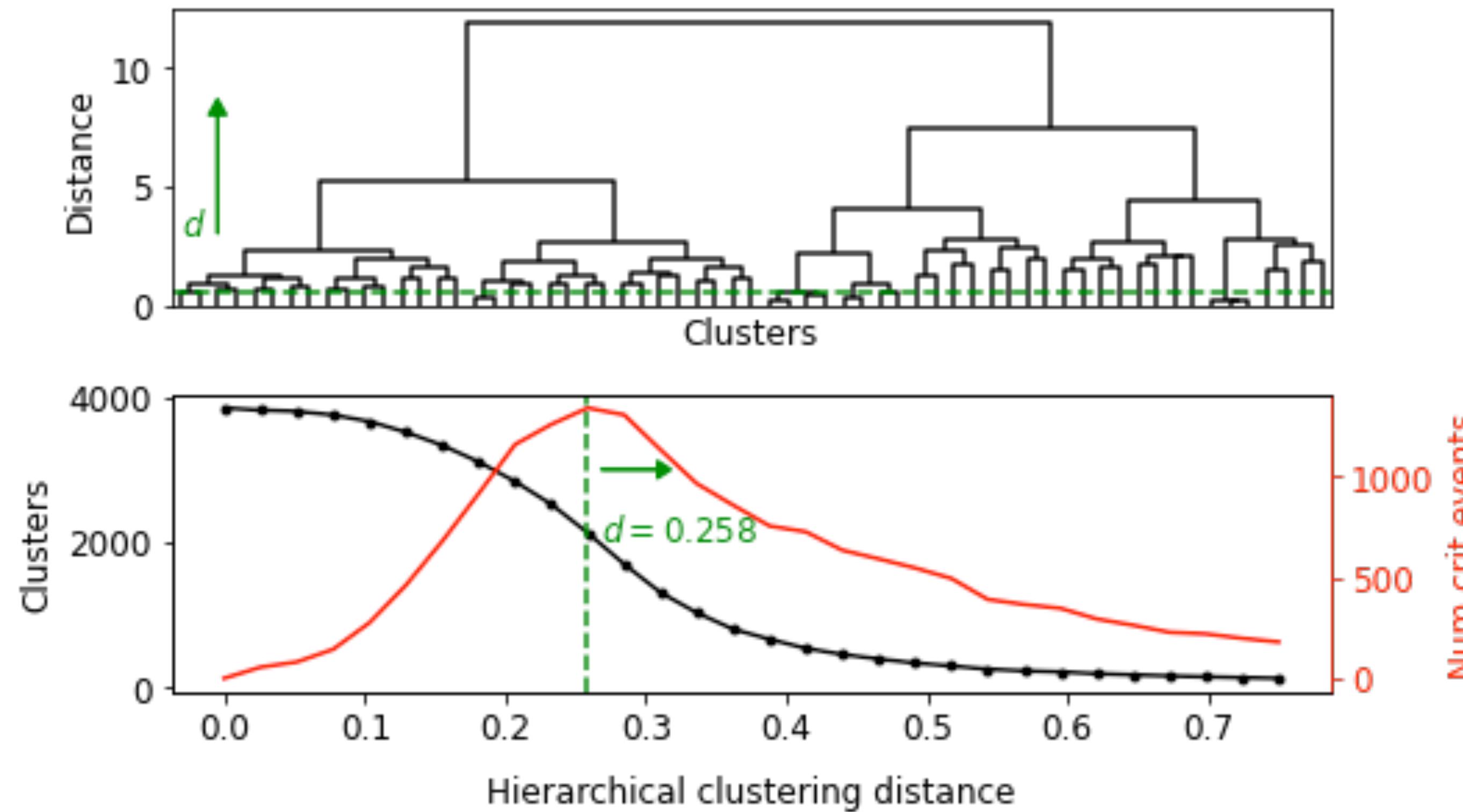


Hierarchical Clustering

w / Ward Linkage



How many clusters: Dendrogram



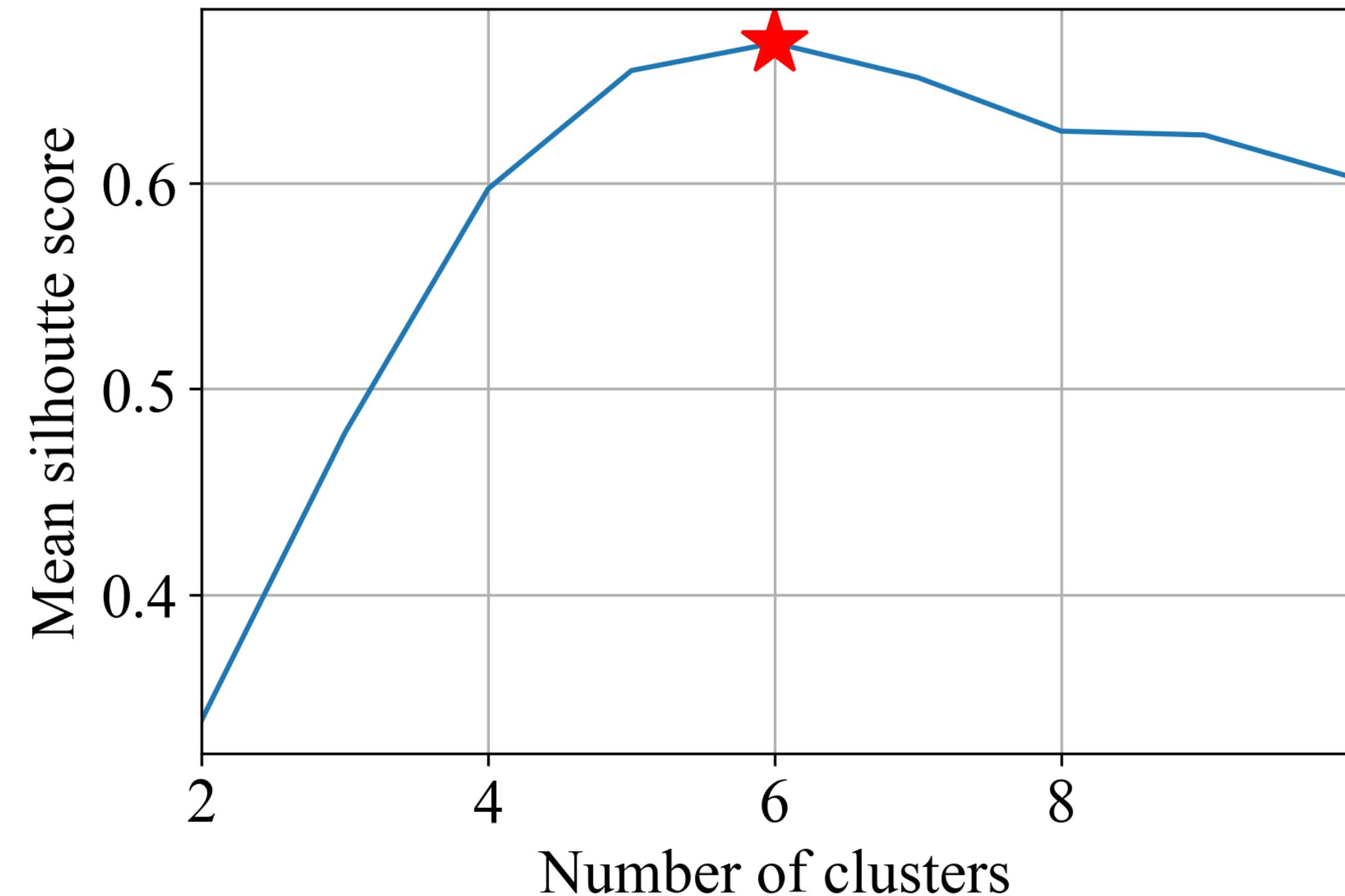
How many clusters: Silhouette Scores (SiS)

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

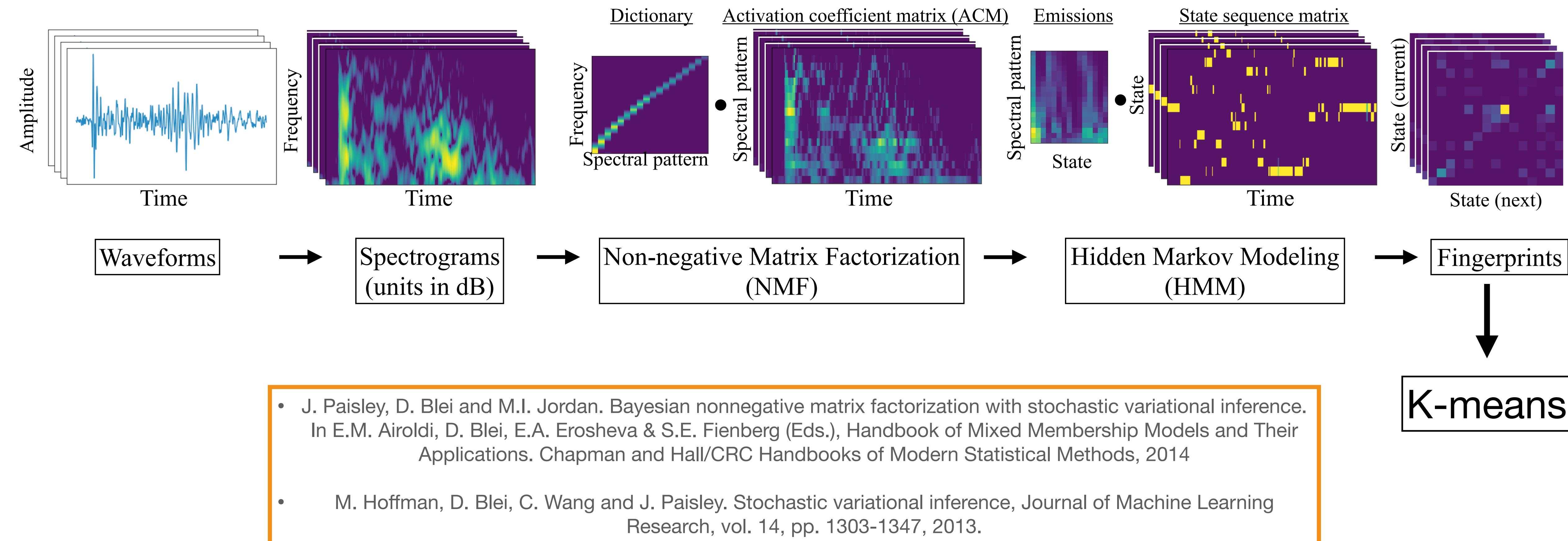
- Unsupervised ML infers patterns in large data sets without the need for prior training labels
- Feature extraction and dimensionality reduction can aid clustering and interpretation
 - *Examples: Nonnegative matrix factorization (NMF), K-means clustering, hierarchical clustering*

Outline

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

Spectral Unsupervised Feature Extraction (SpecUFEx)

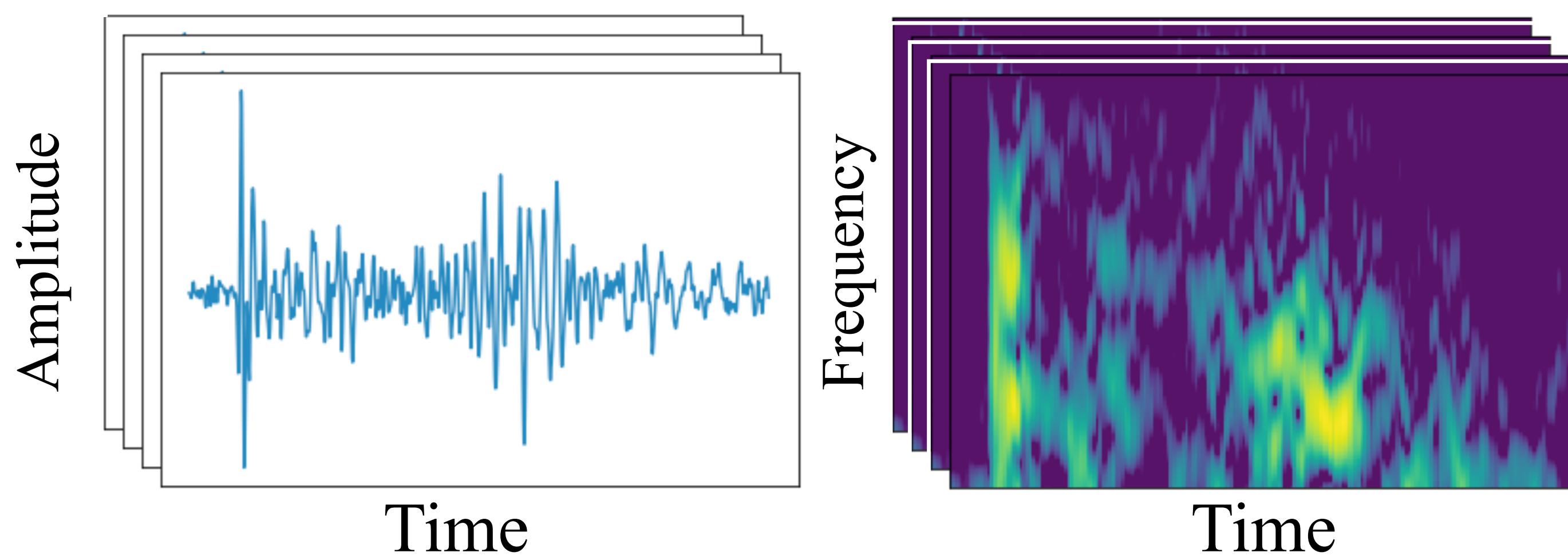
Holtzman et al., 2018; Sci Adv



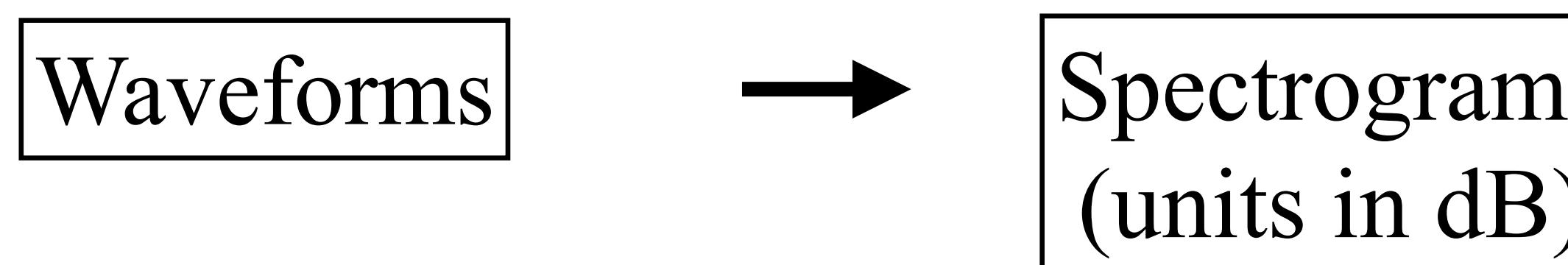
Spectral Unsupervised Feature Extraction (SpecUFEx)

Holtzman et al., 2018; Sci Adv

Generate spectrogram



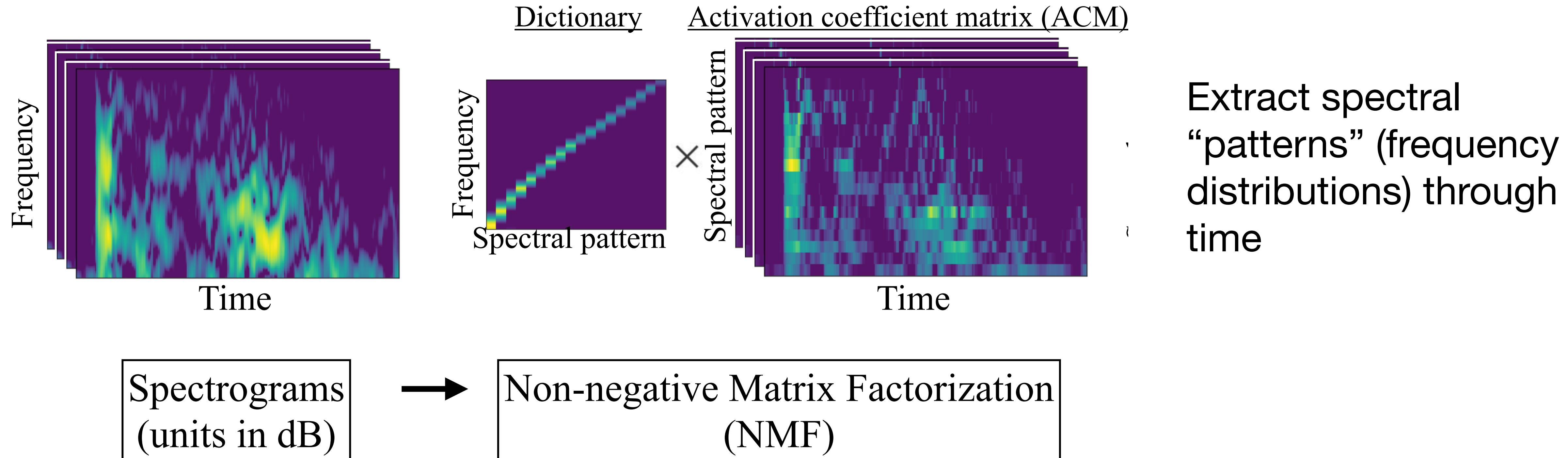
Ground amplitude through time ->
Frequency content through time



Spectral Unsupervised Feature Extraction (SpecUFEx)

Holtzman et al., 2018; Sci Adv

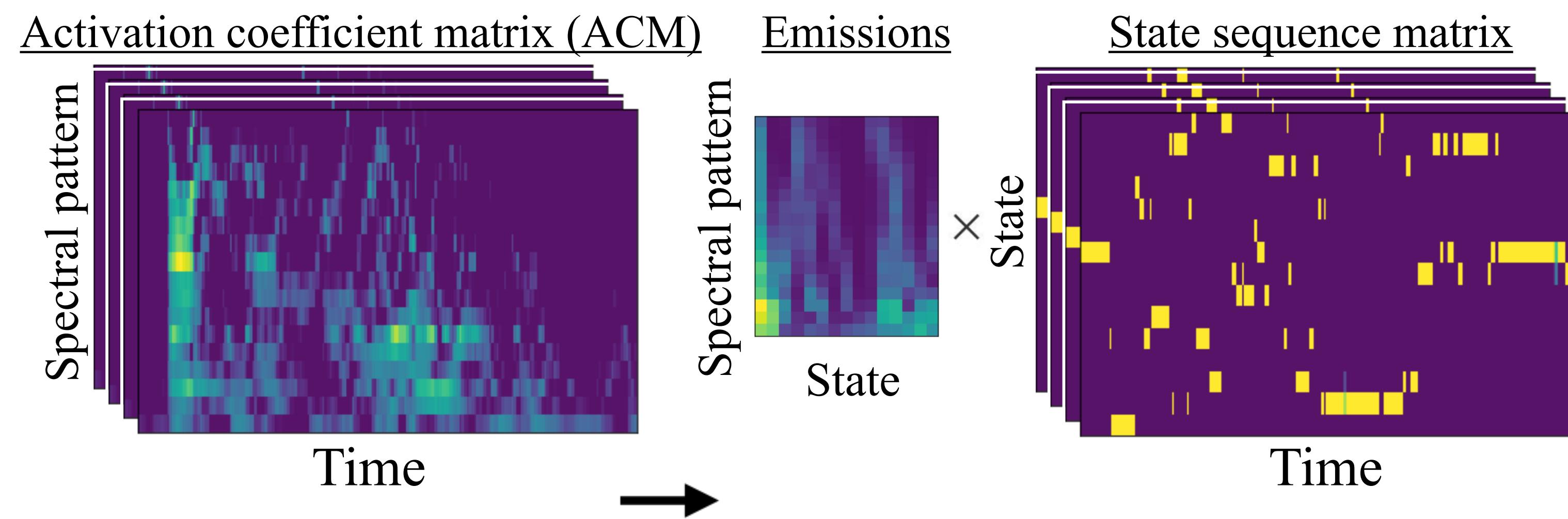
Non-negative Matrix Factorization



Spectral Unsupervised Feature Extraction (SpecUFEx)

Holtzman et al., 2018; Sci Adv

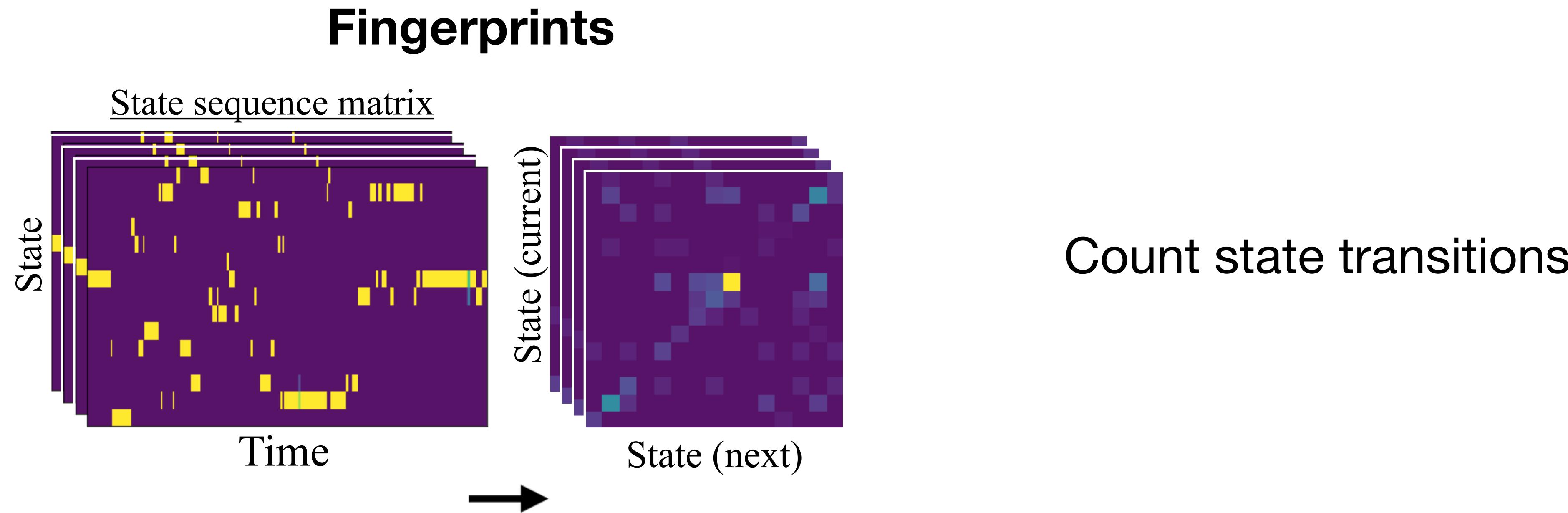
Hidden Markov Model



Hidden states control concurrent spectral patterns

Spectral Unsupervised Feature Extraction (SpecUFEx)

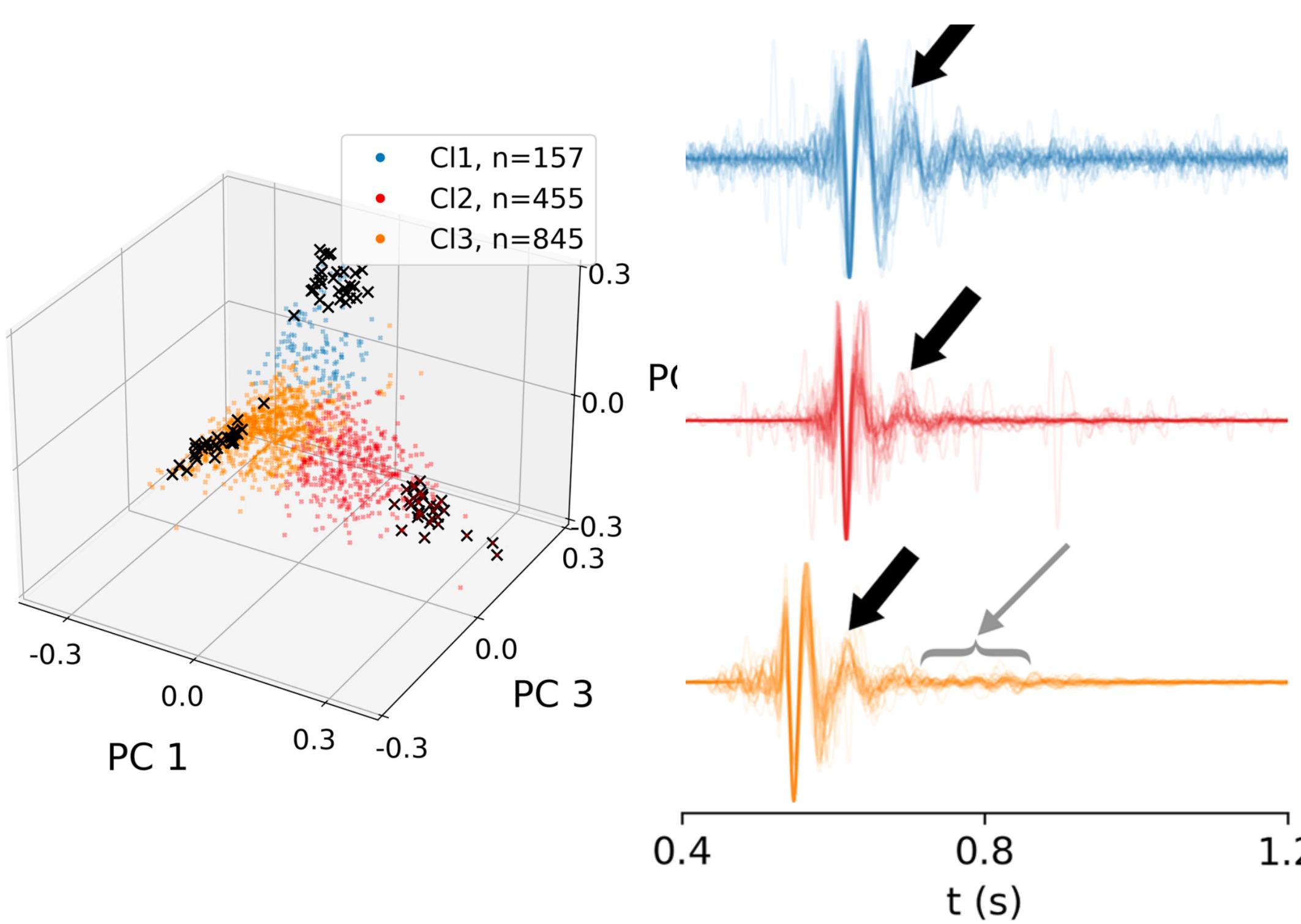
Holtzman et al., 2018; Sci Adv



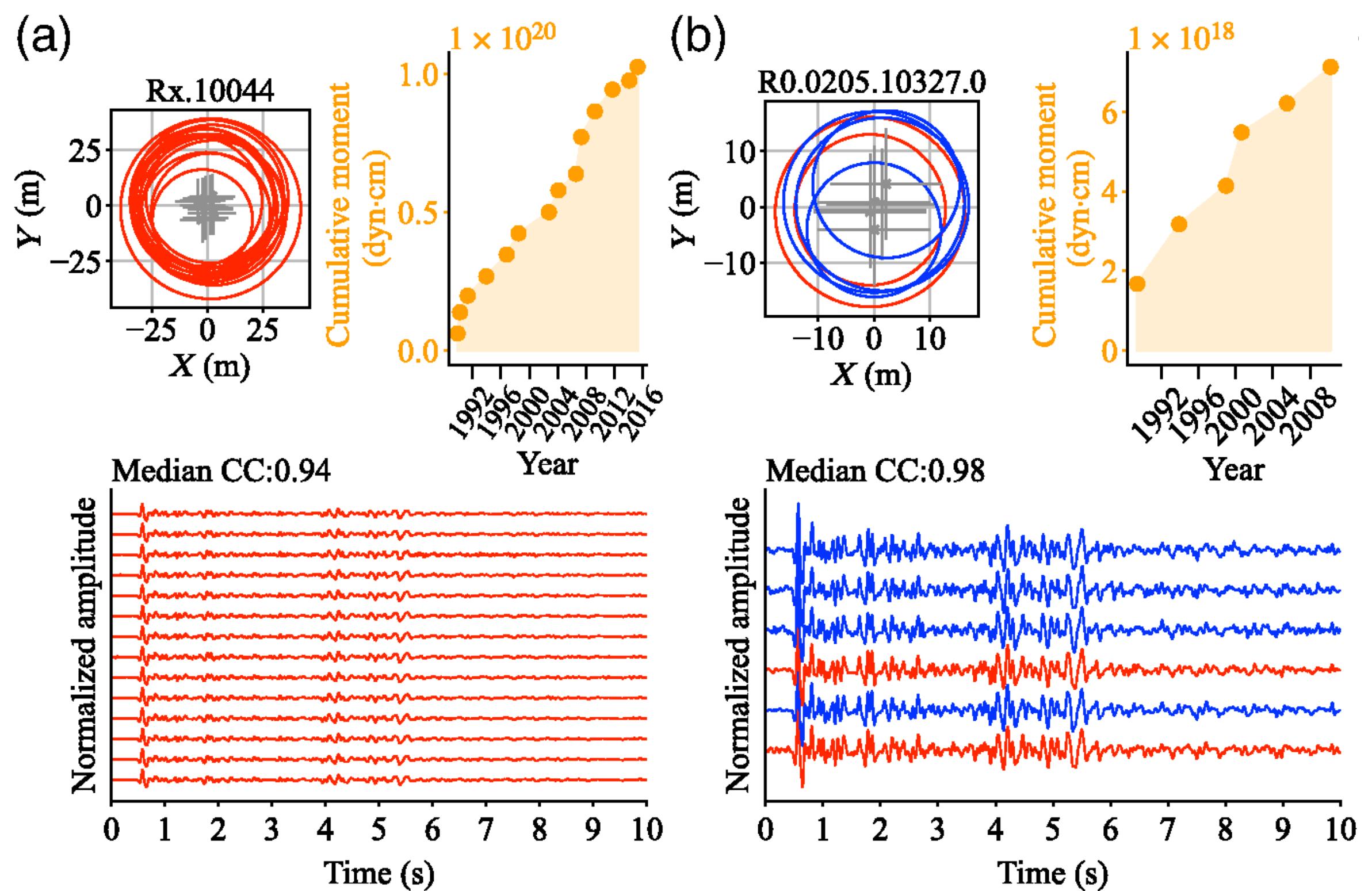
Examples of SpecUFEx

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*

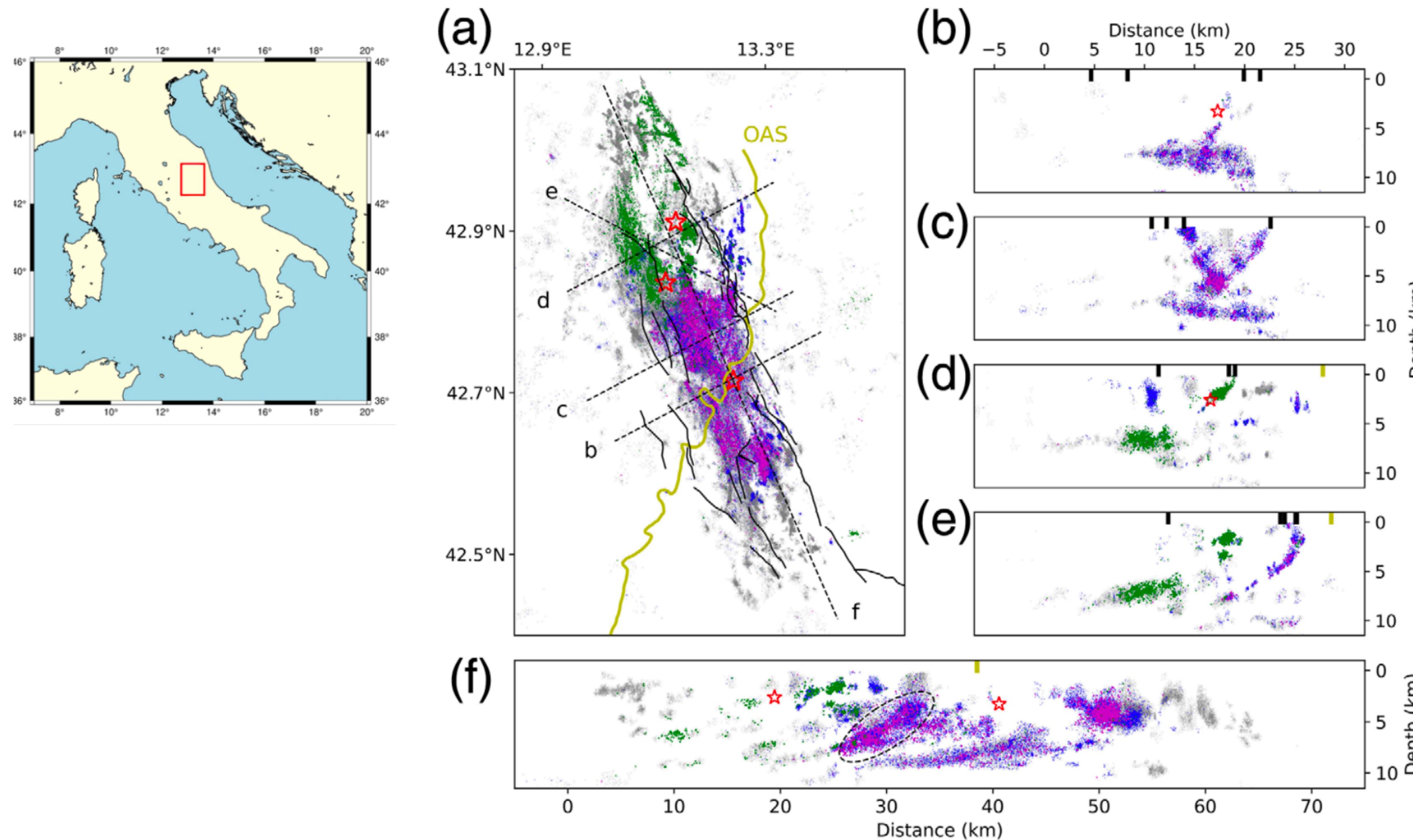


Detecting Repeating Earthquakes on the San Andreas Fault with Unsupervised Machine Learning of Spectrograms (2023) Sawi T., Waldhauser F., Holtzman, B.K., Groebner, N.; . *The Seismic Record*



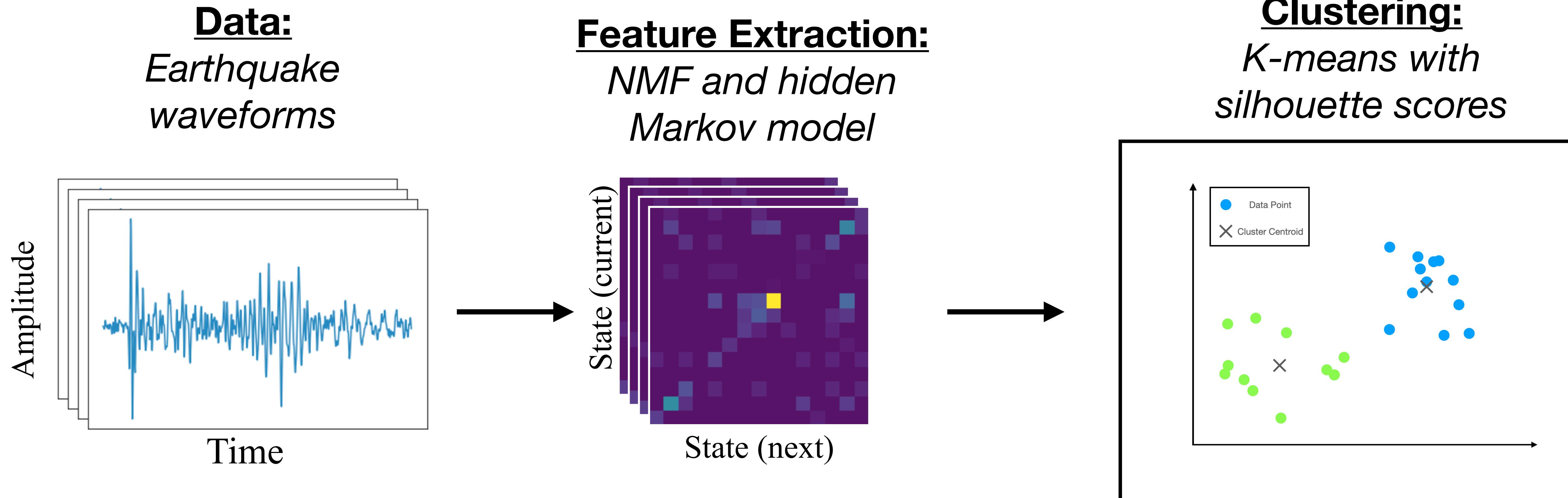
Amatrice-SpecUFEx Tutorial

(subset = 6 days, ~1000 earthquakes)



Tan et al., 2021; The Seismic Record

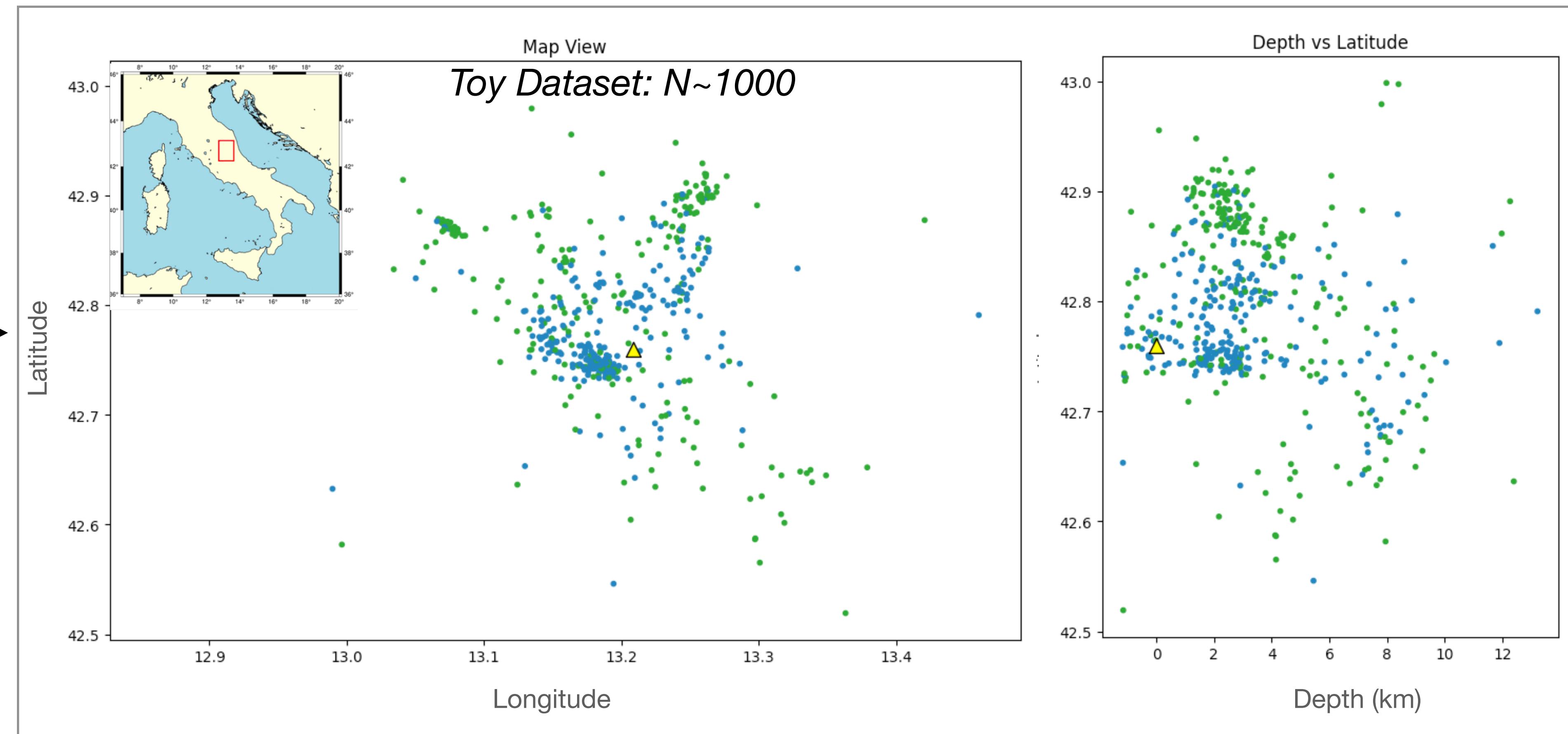
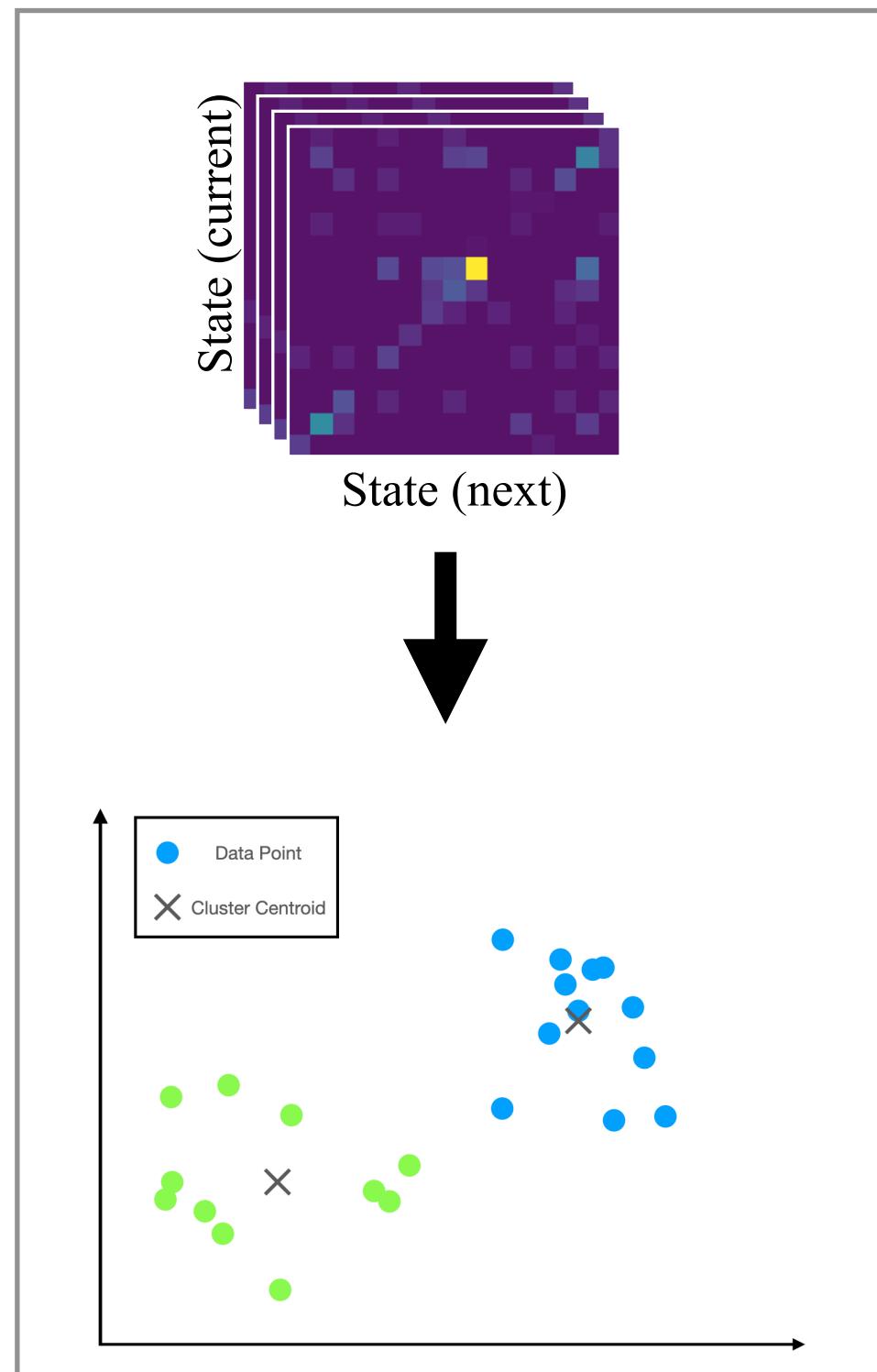
Amatrice-SpecUFEx Tutorial



SpecUFEx

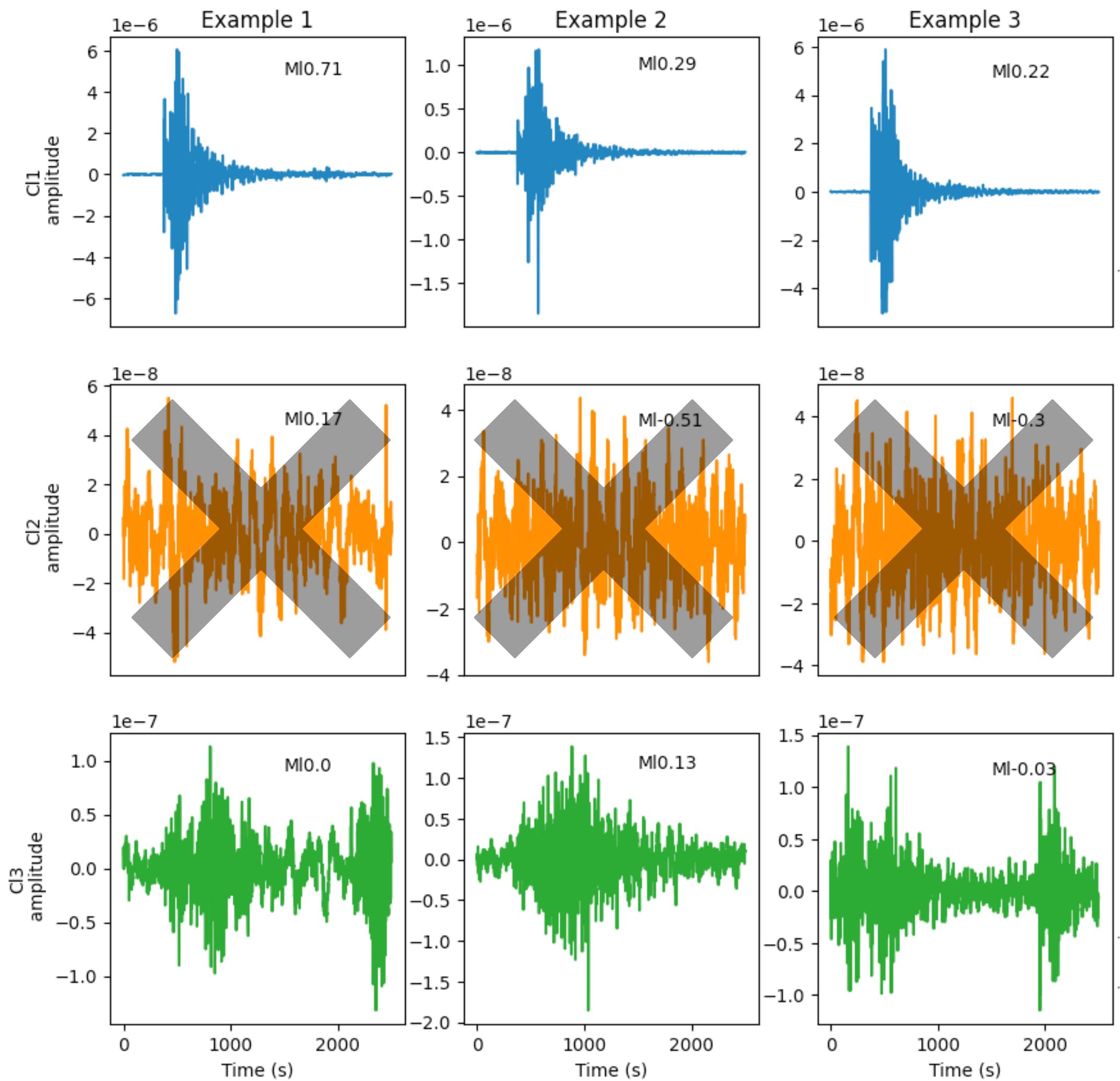
+

K-means clustering



Results for K=3, with 2nd cluster removed

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

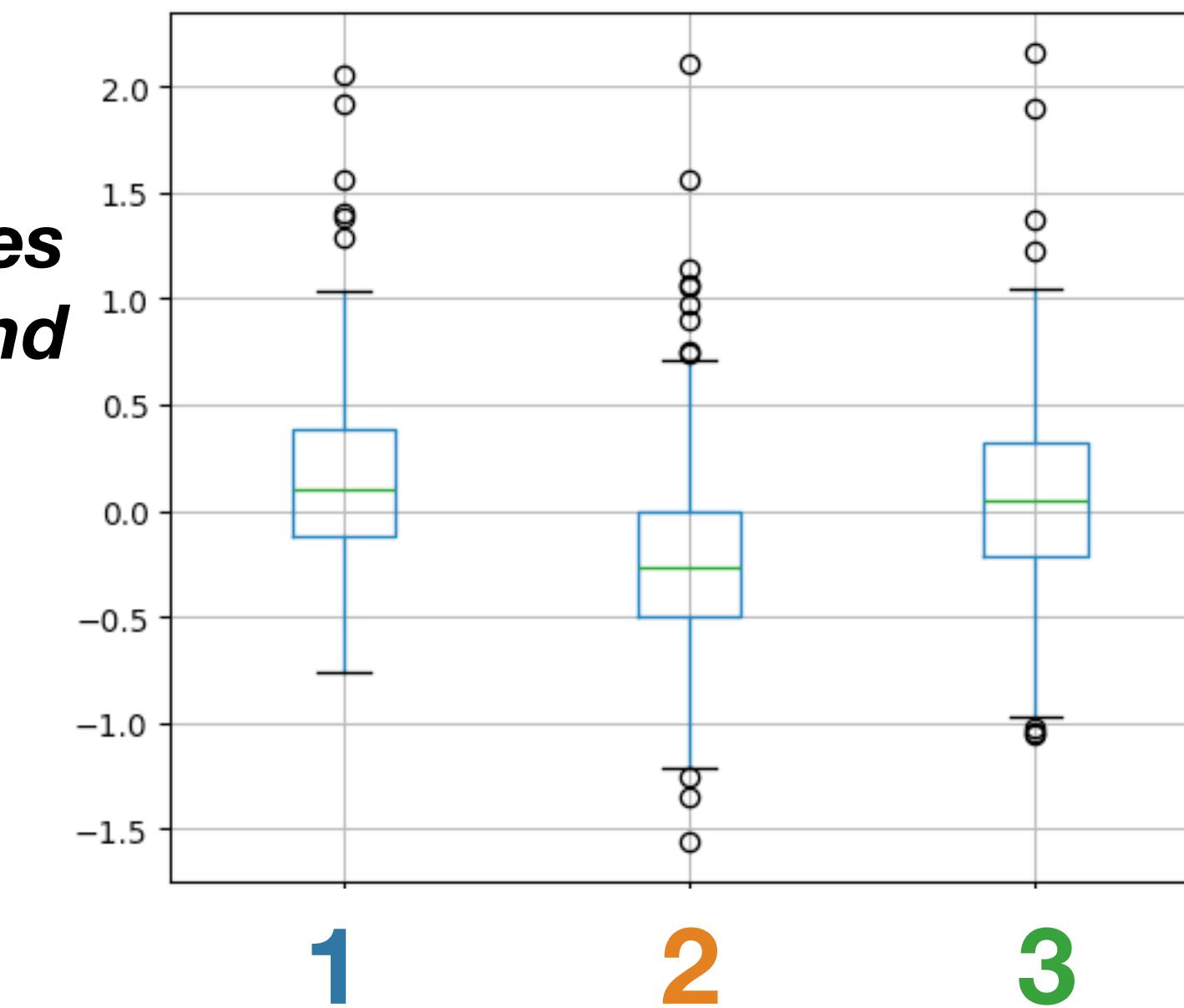


Amatrice-SpecUFEx Tutorial

Estimated magnitudes

**Similar
magnitudes
for blue and
green**

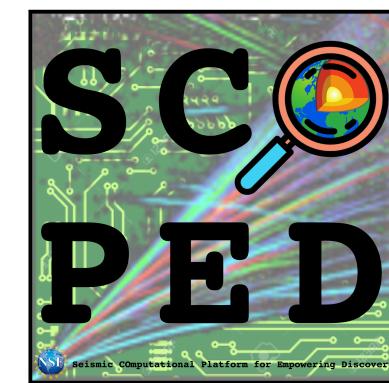
**Noisy waveforms
- bad picks?**



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

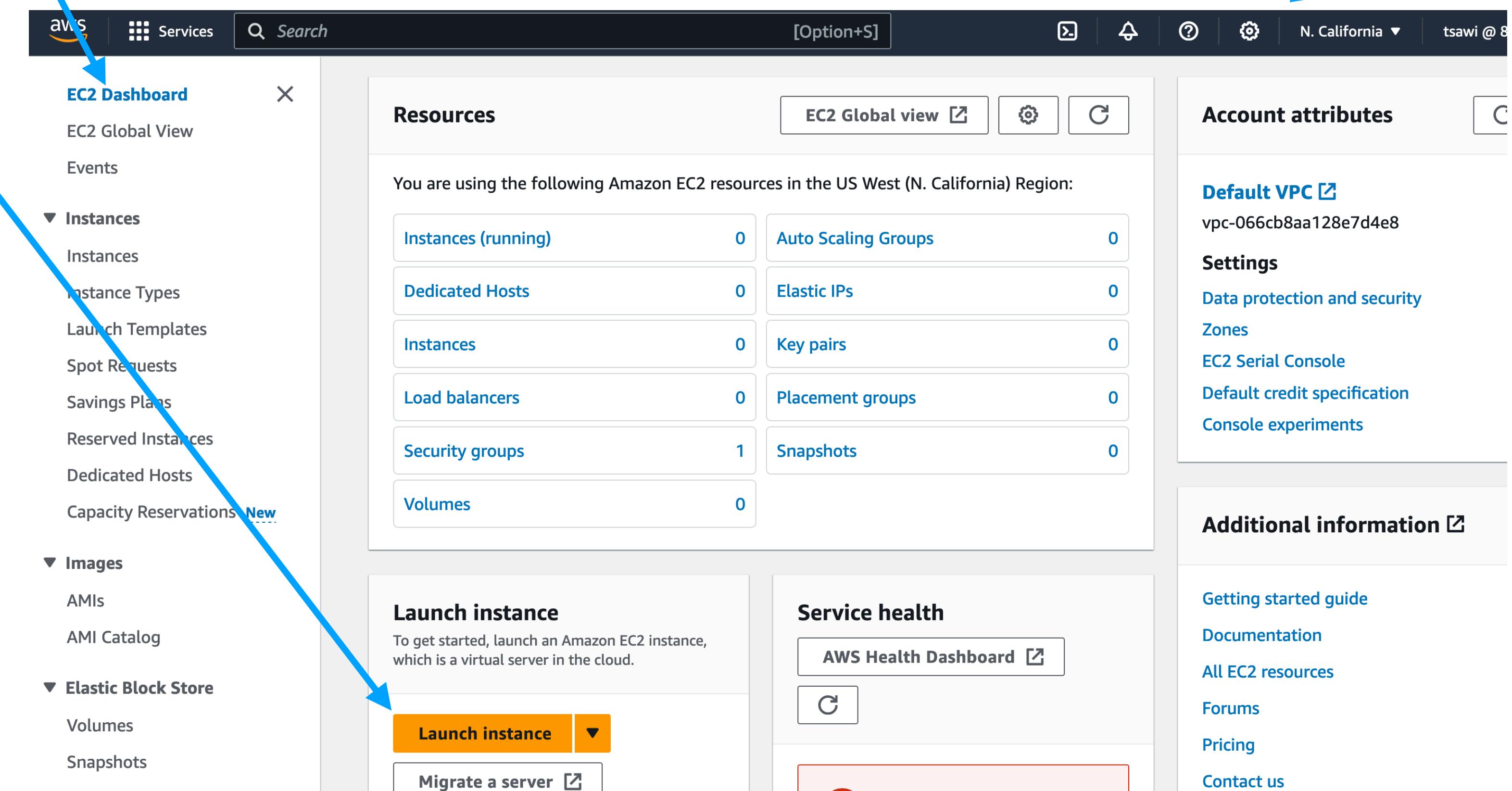
Theresa Sawi, Nate Groebner, Felix Waldhauser, Kaiwen Wang, Yen-Joe Tan

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 SpecUFEEx/2016 Amatrice sequence

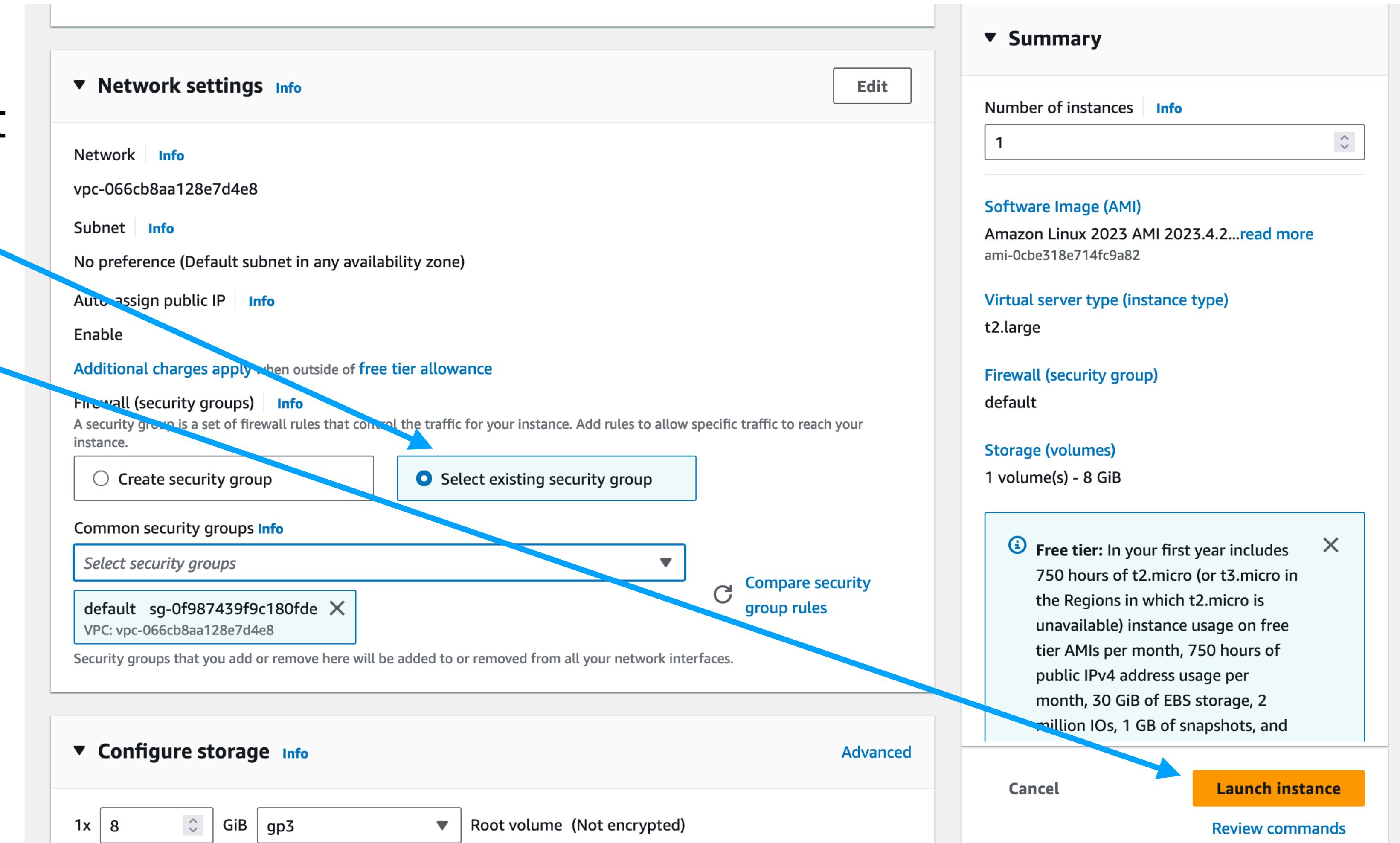
Using AWS Cloud

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



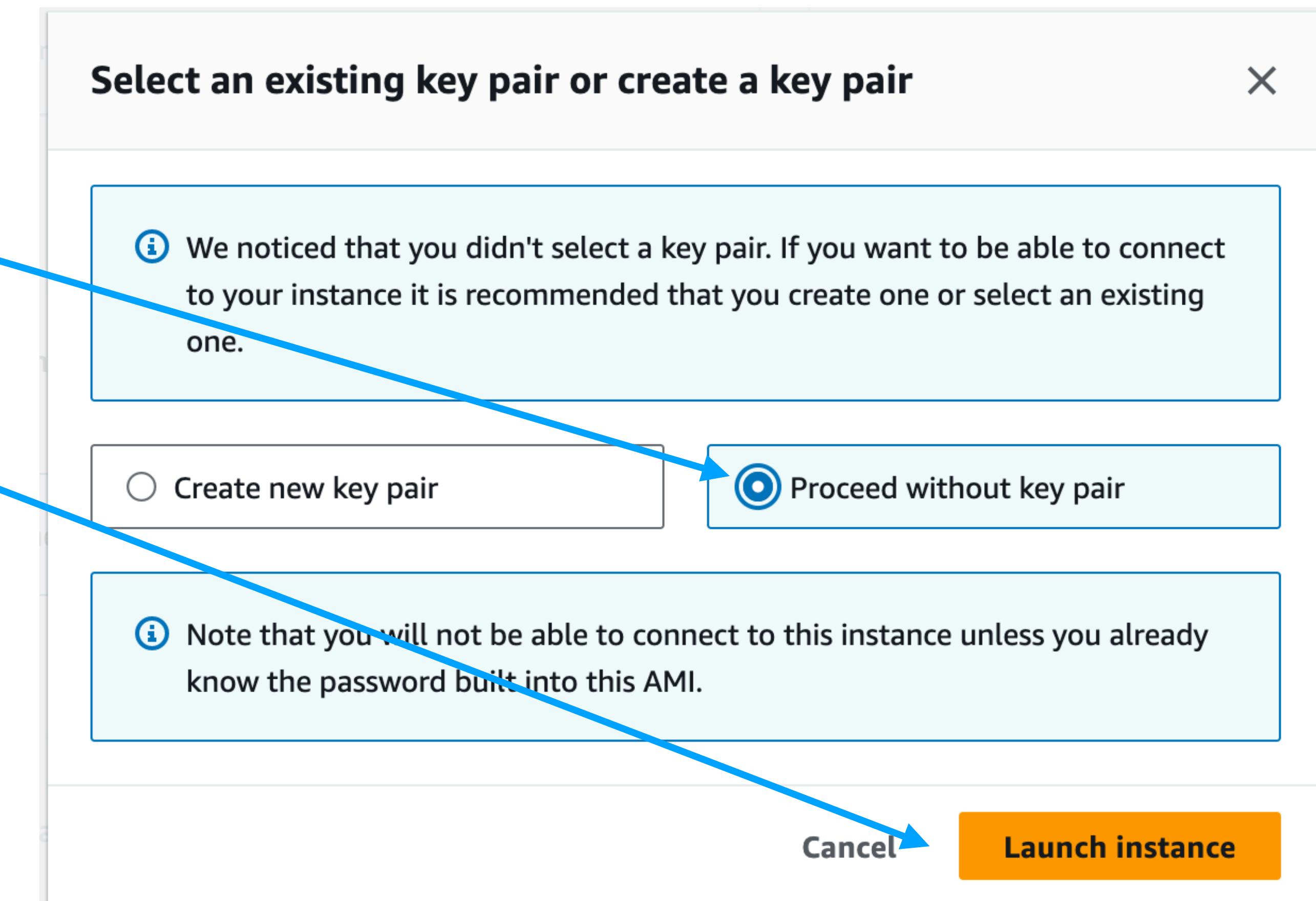
Using AWS Cloud

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



Using AWS Cloud

1. Select “proceed without key pair”
2. 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

Instance summary for i-031c536773da6554b (testssawi)

C Connect Instance state ▾ Actions ▾

Updated less than a minute ago

Instance ID <input type="button"/> i-031c536773da6554b (testssawi)	Public IPv4 address <input type="button"/> 54.176.62.87 open address
IPv6 address -	Instance state 🕒 Running

EC2 > Instances > Launch an instance

Success
Successfully initiated launch of instance (i-031c536773da6554b)

▶ Launch log

Next Steps

What would you like to do next with this instance, for example “c”

Instances (1) Info C Connect Instance state ▾ Actions ▾ Launch instance

Find Instance by attribute or tag (case-sensitive) All states ▾

Instance ID = i-031c536773da6554b X Clear filters

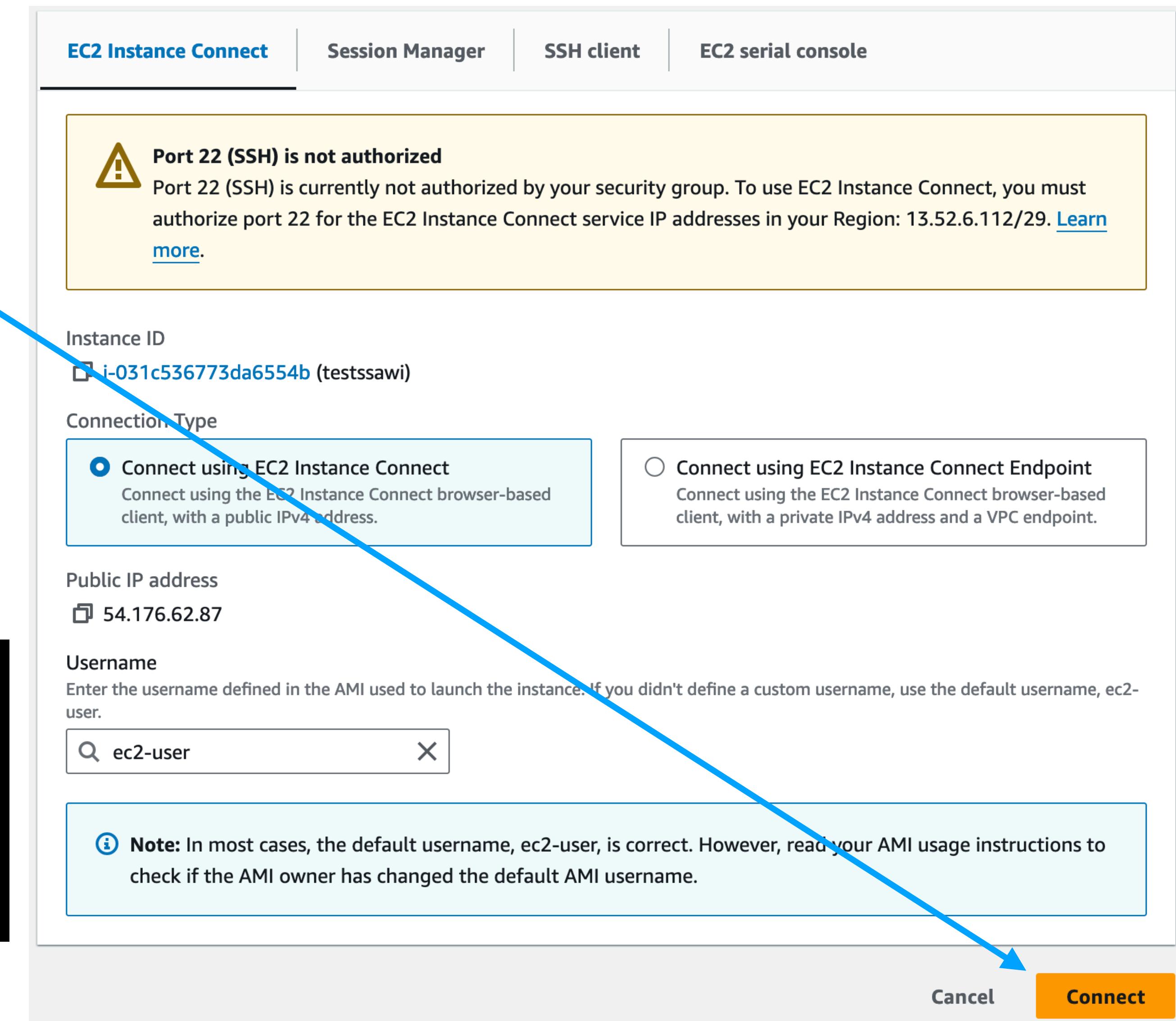
<input type="checkbox"/>	Name <input type="button"/>	Instance ID	Instance state	Instance type	Status check
<input type="checkbox"/>	testssawi	i-031c536773da6554b	🕒 Running <input type="button"/> <input type="button"/>	t2.large	⌚ Initializing

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
--       V~'__->
---           /
--..-./_/
-/m/ '
[ec2-user@ip-172-31-23-105 ~]$ sudo yum install -y git docker
wget 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
```



Tutorial Notebook (static)

sawilabs.com/cloud.html

These Slides

sawilabs.com/slides.pdf

Download Docker image locally

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