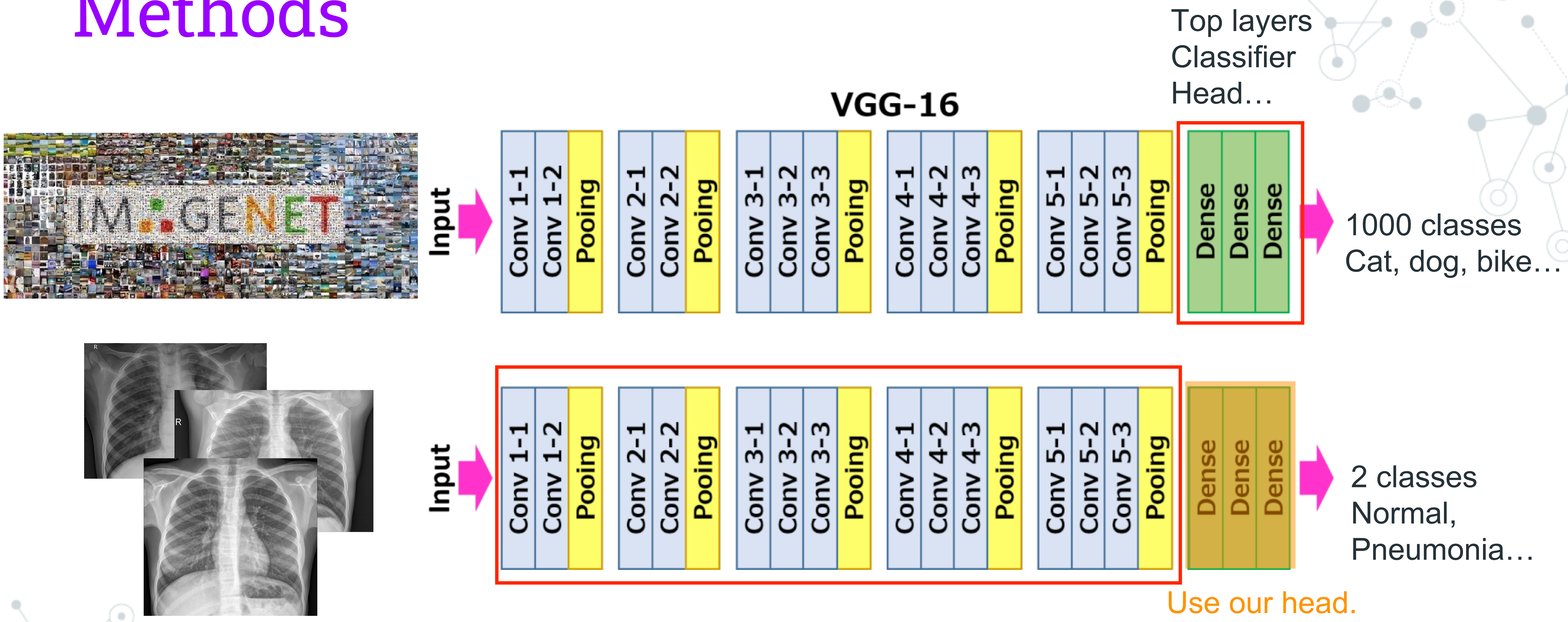
Lab 6: Unsupervised Learning

11210IPT 553000

Deep Learning in Biomedical Optical Imaging

2023/10/30

Methods



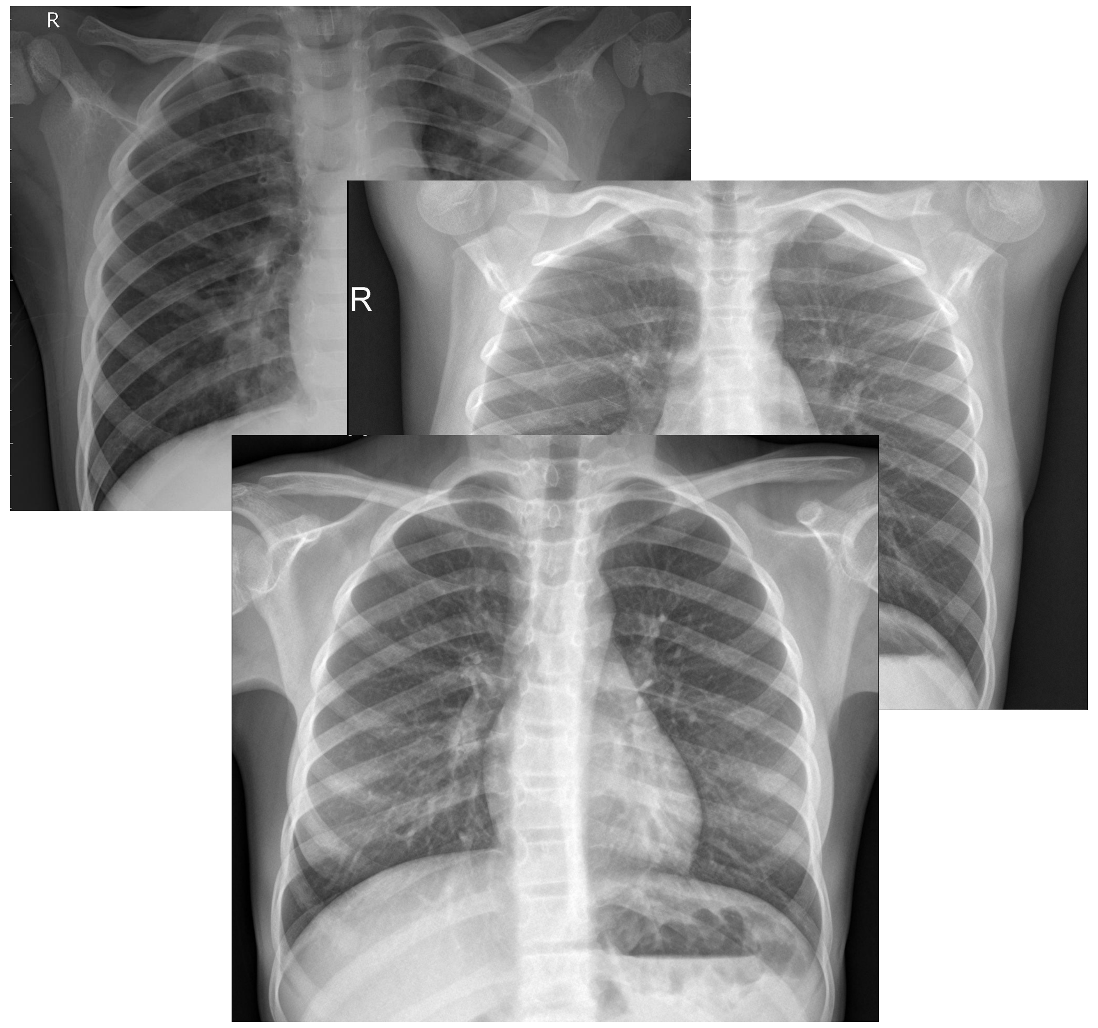
Use pre-trained model as a **feature extractor**.

Transfusion: Understanding Transfer Learning for Medical Imaging

- A performance evaluation on two large scale medical imaging tasks shows that surprisingly, **transfer offers little benefit** to performance, and simple, lightweight models can perform comparably to ImageNet architectures.
- Some of the differences from transfer learning are due to the **over-parametrization** of standard models rather than complex feature reuse.
  - The large, standard ImageNet models do not change significantly through the fine-tuning process.

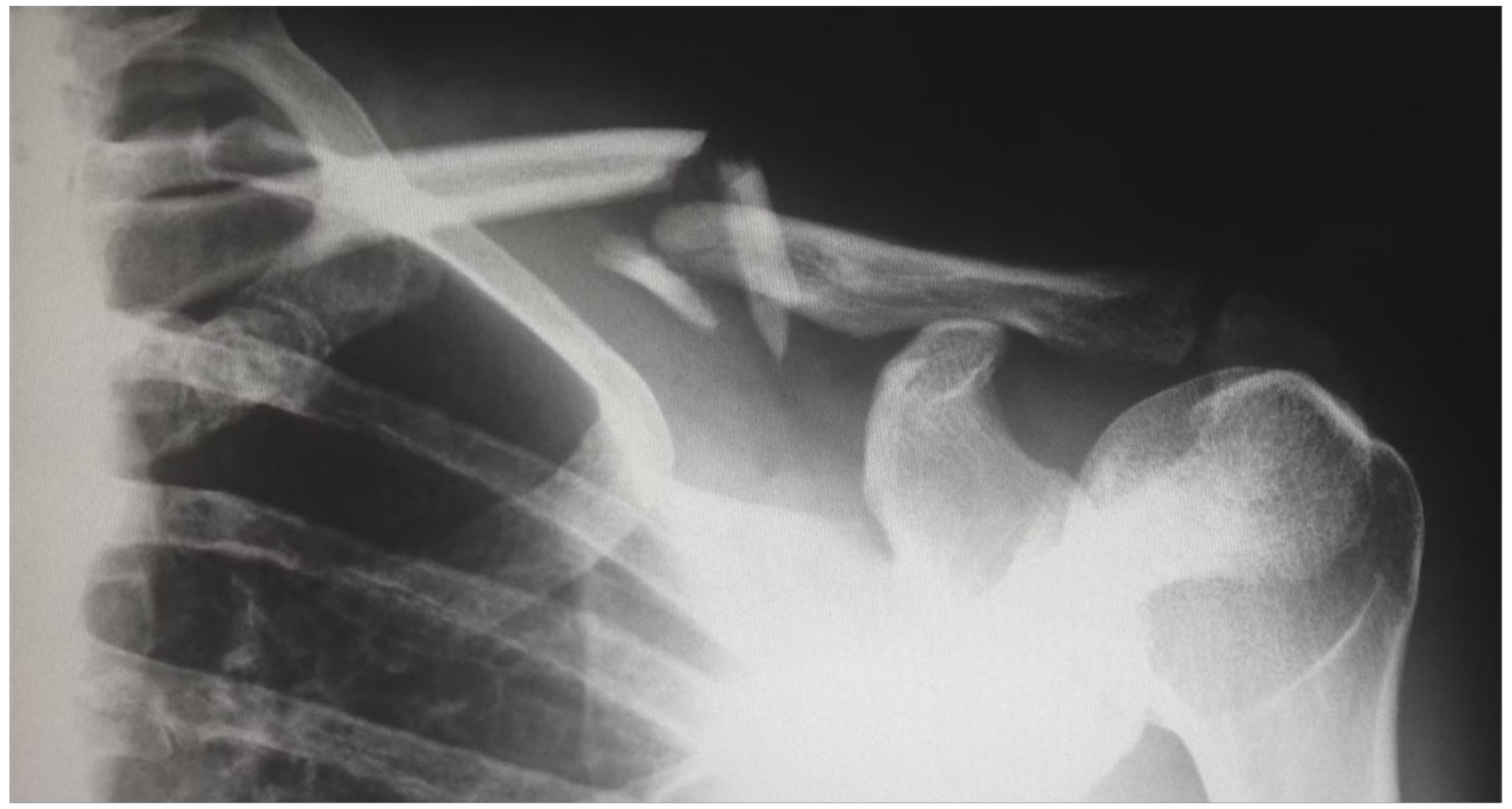
Raghu, Maithra, et al. "Transfusion: Understanding transfer learning for medical imaging." *NeruIPS* (2019).

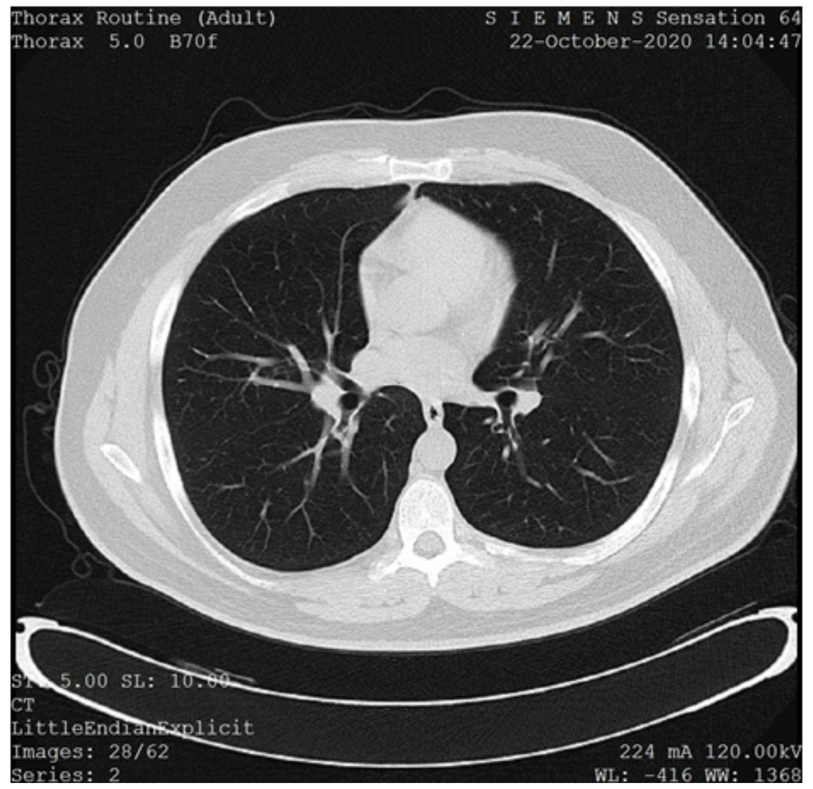
Difference between these two datasets



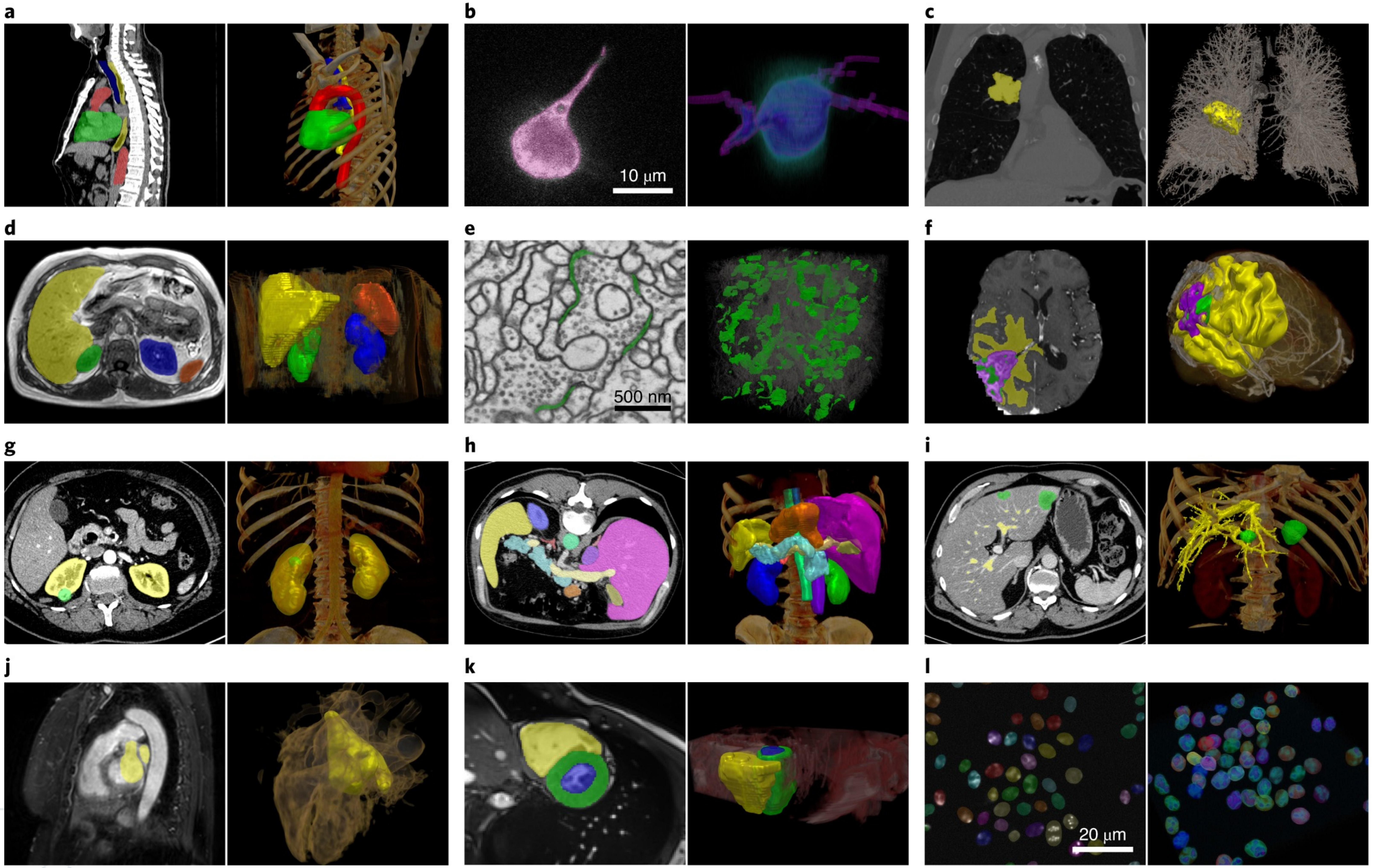
The ImageNet Dataset

Why not just pre-train feature extractors on **biomedical datasets**?





Because labeling is **very expensive** in this domain



https://www.nature.com/articles/s41592-020-01008-z

How to train deep neural networks with large amounts of

unlabeled data?

Let’s get started with

unsupervised Learning.

Outlines

- What is unsupervised learning
- Methods and applications
- Unsupervised representation (feature) learning

Outlines

- What is unsupervised learning
- Methods and applications
- Unsupervised representation (feature) learning

Unsupervised learning is a type of machine learning where algorithms learn from data **without explicit labels**.

- Learn from **paired** data: **supervised** learning

$$(x_1, y_1), \ldots, (x_n, y_n)$$

- Learn from **unpaired** data: **unsupervised** learning

$$(x_1), \ldots, (x_n)$$

Unsupervised learning tries to learn the

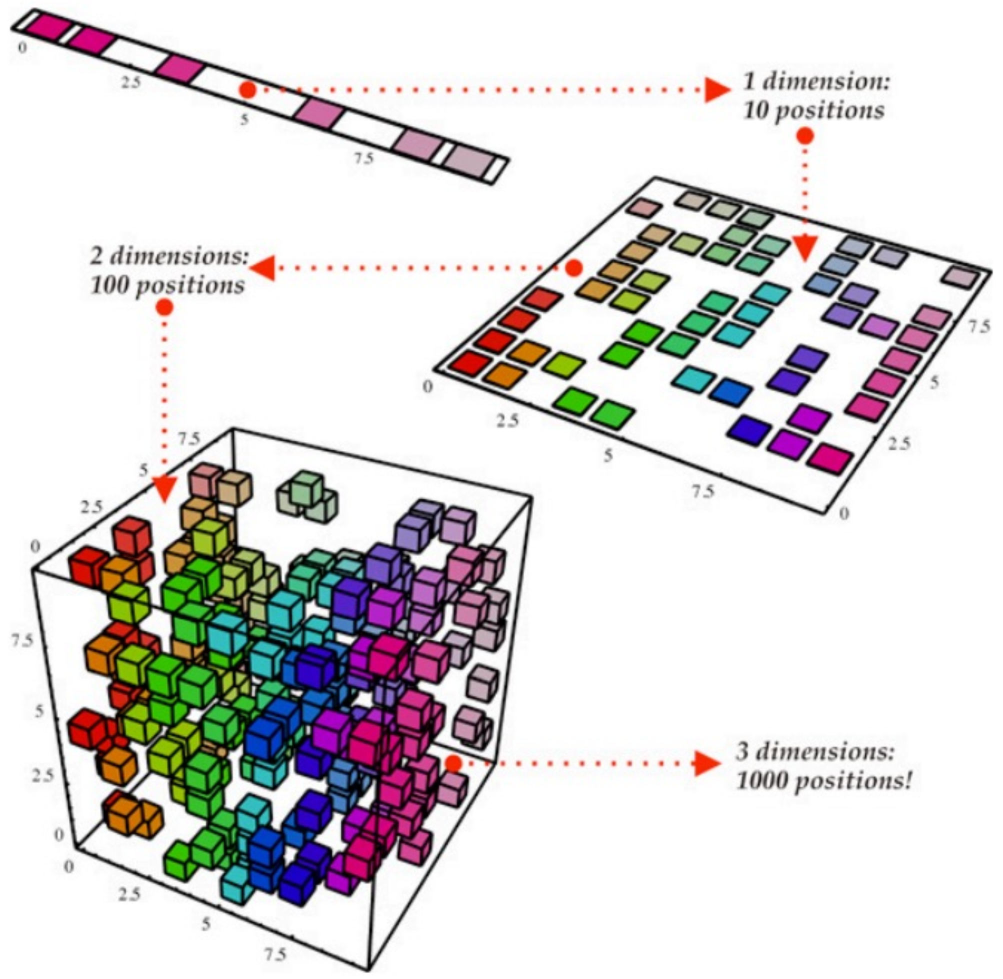
underlying structure

from the data without any guidance.

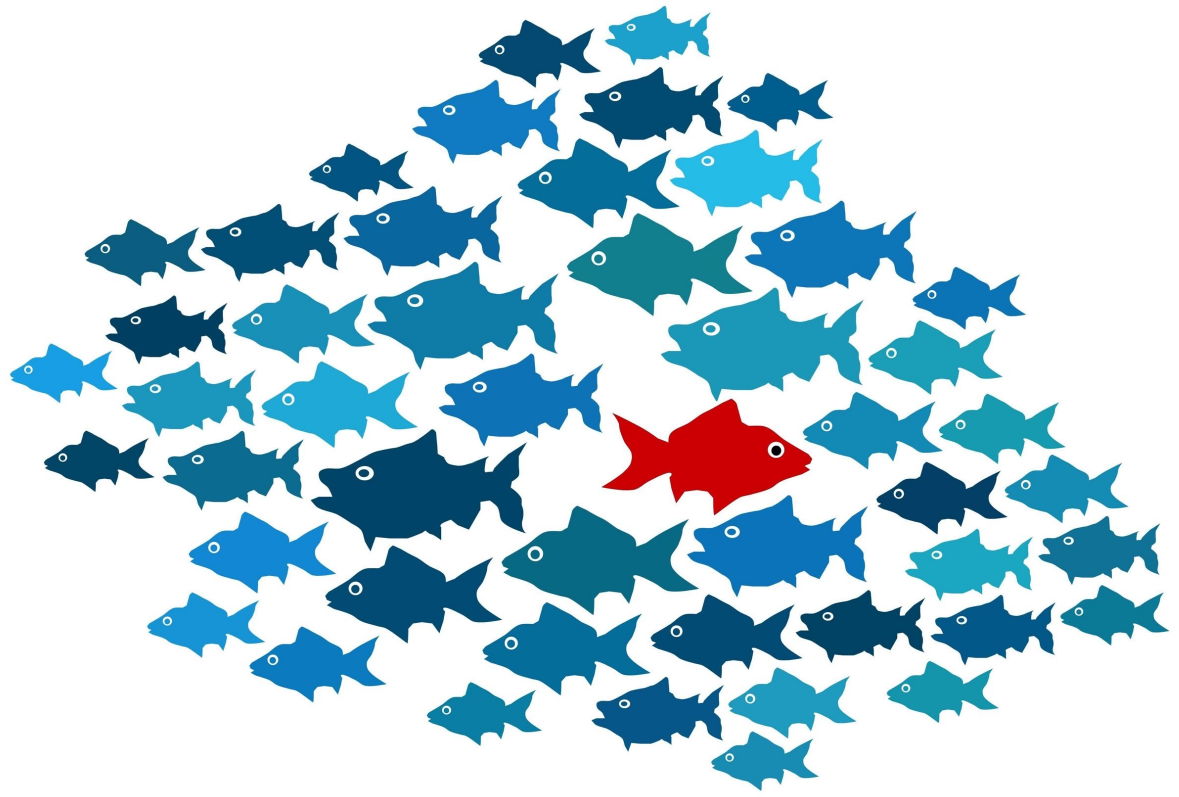
Some common tasks and techniques in unsupervised learning

Clustering

Dimension Reduction



Anomaly Detection



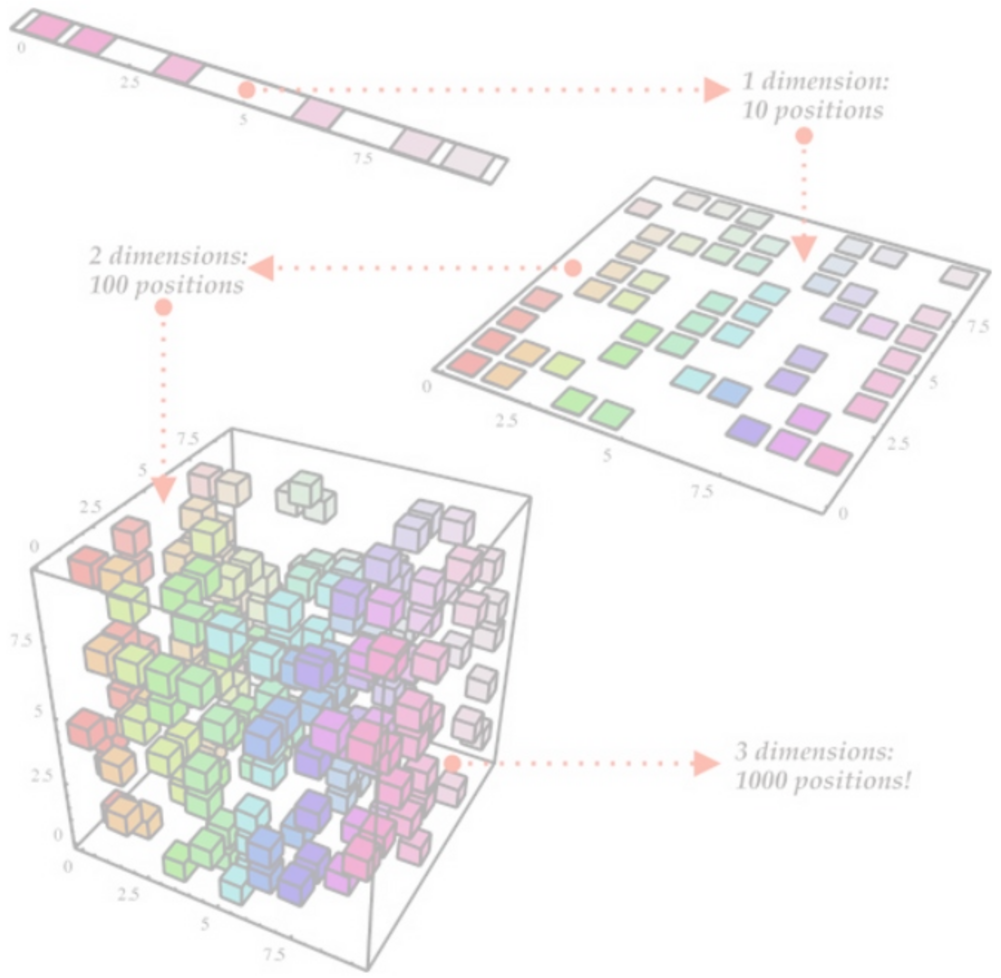
Outlines

- What is unsupervised learning
- Methods and applications
- Unsupervised representation (feature) learning

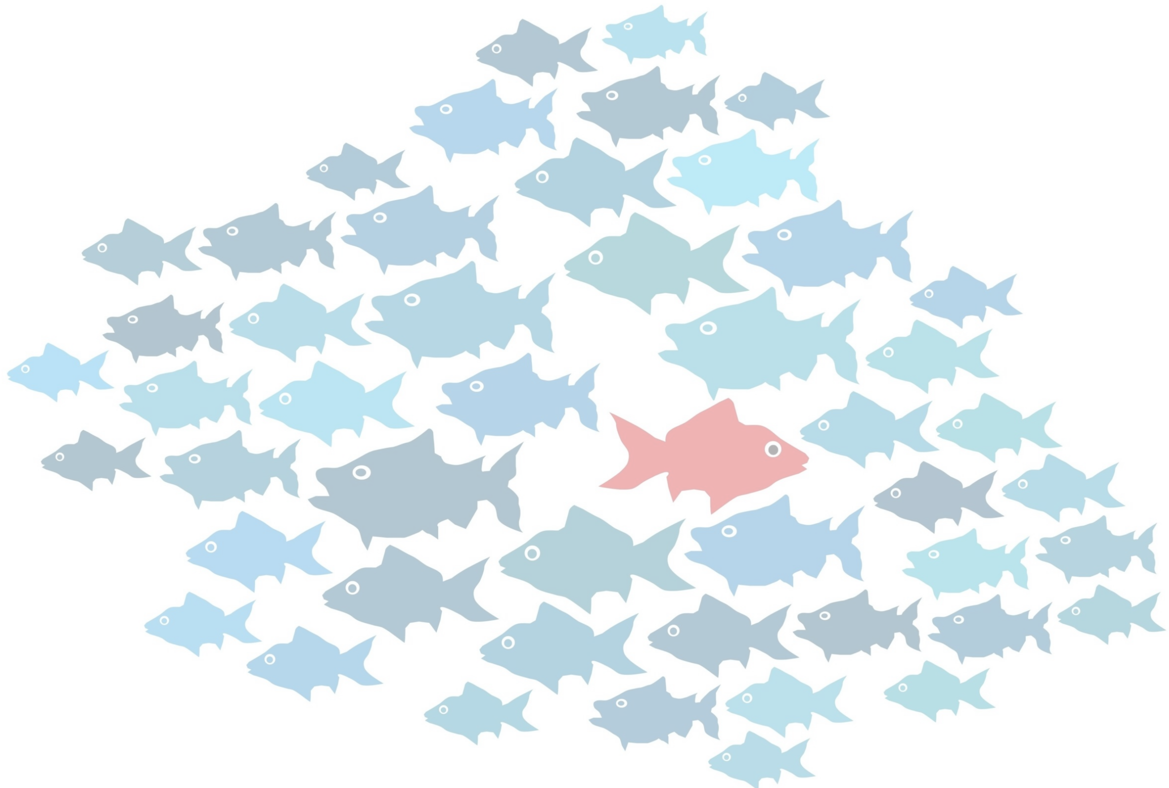
Some common tasks and techniques in unsupervised learning

Clustering

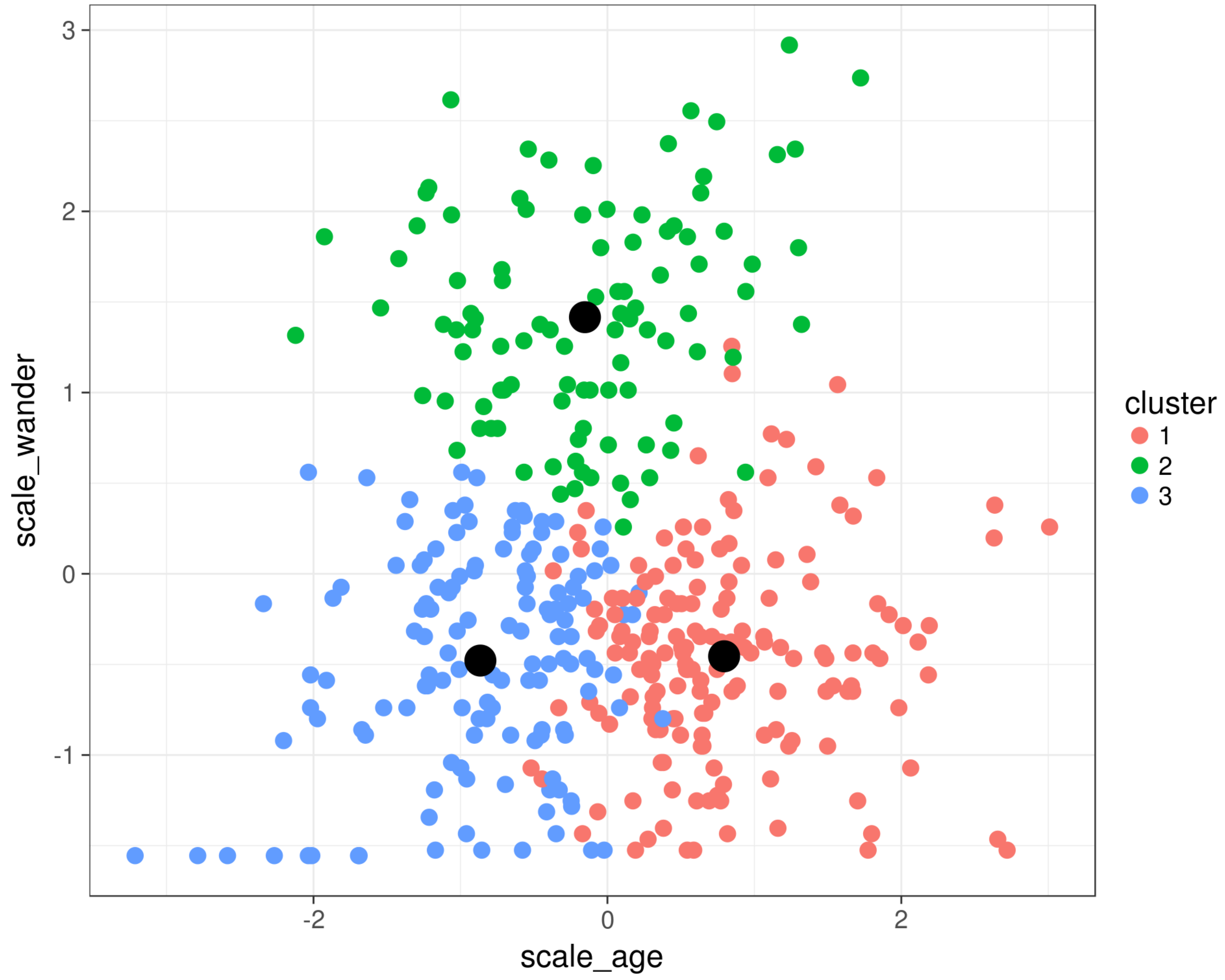
Dimension Reduction

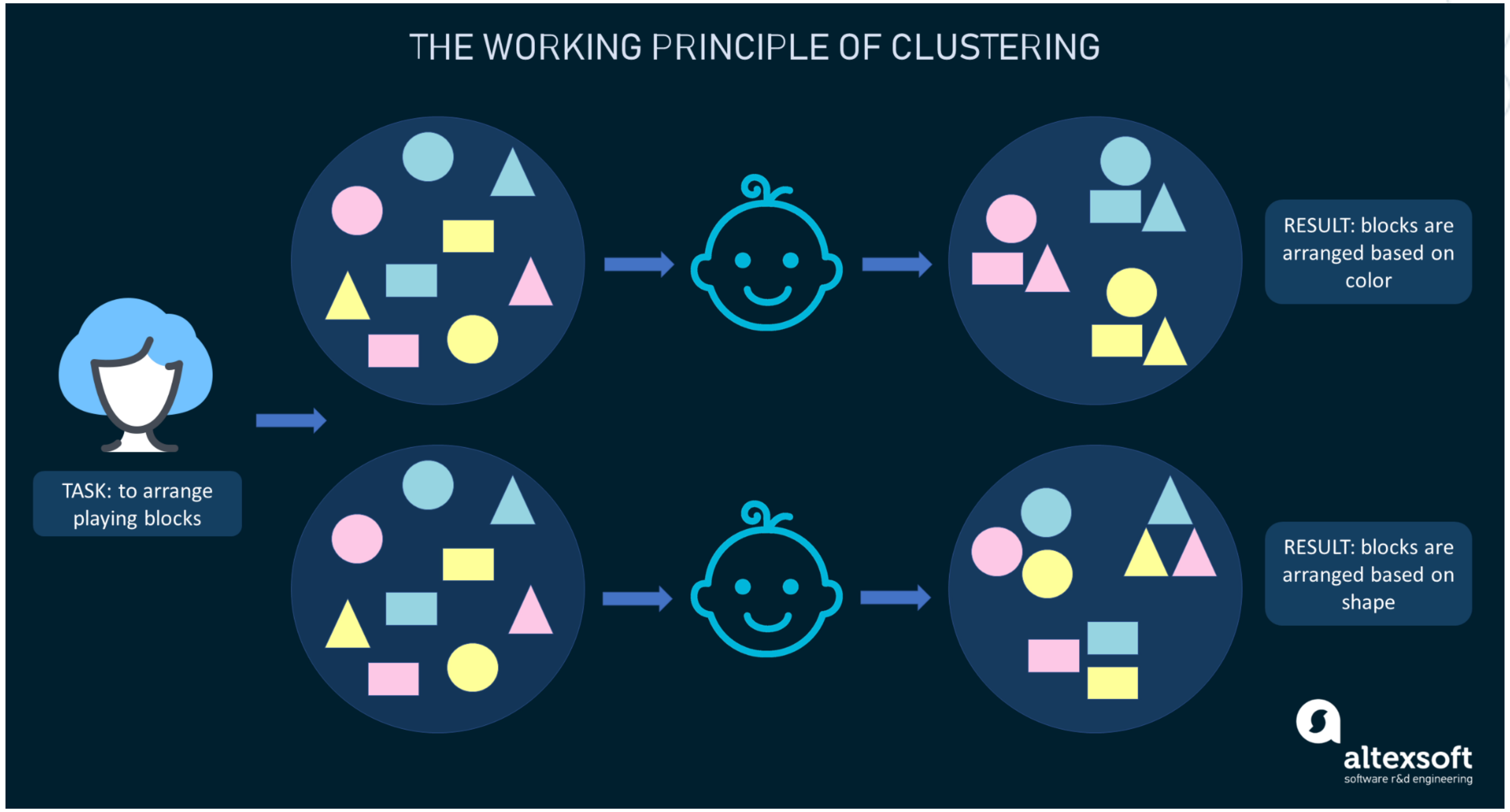


Anomaly Detection

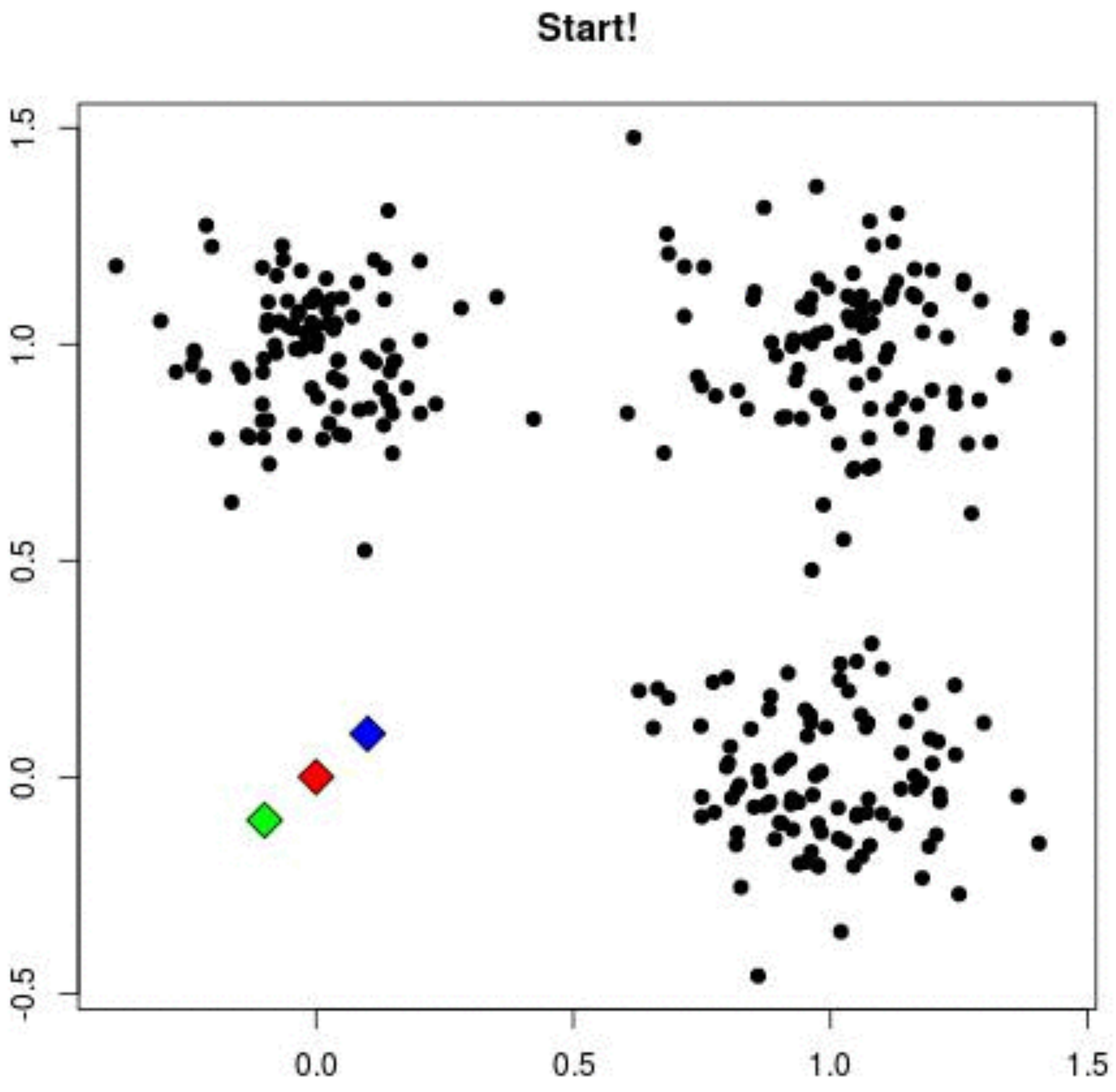


Clustering: group data points into separate clusters or groups based on **similarity**



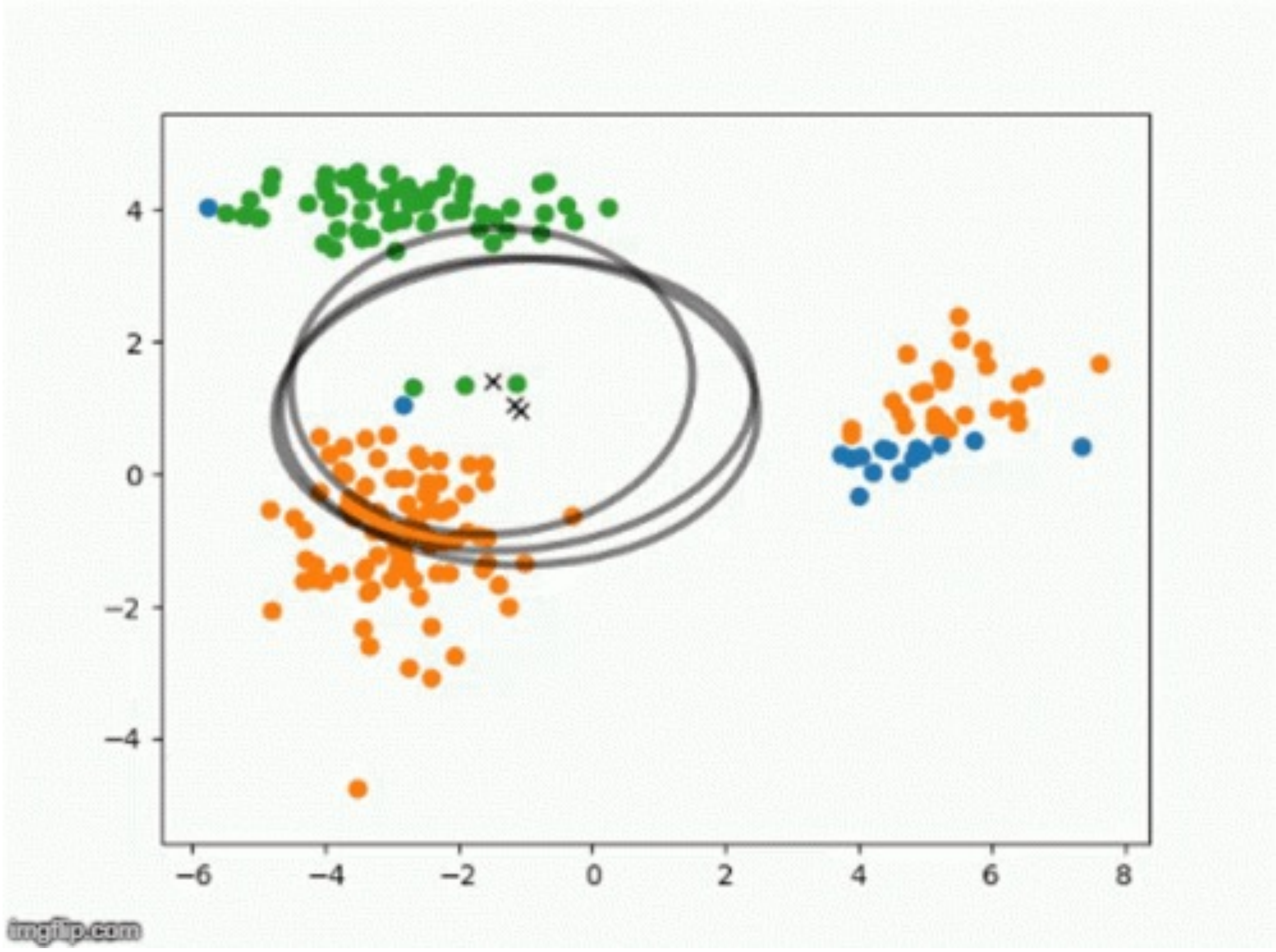


KMeans



1. Specify the number of clusters “K”
2. Randomly initialize the cluster centers
3. Assign each data point to the **closed cluster center**.
4. Recompute the cluster’s centers as the mean of all data in that cluster
5. Repeat 3 and 4 until the cluster assignment stops changing or the maximum iteration is reached.

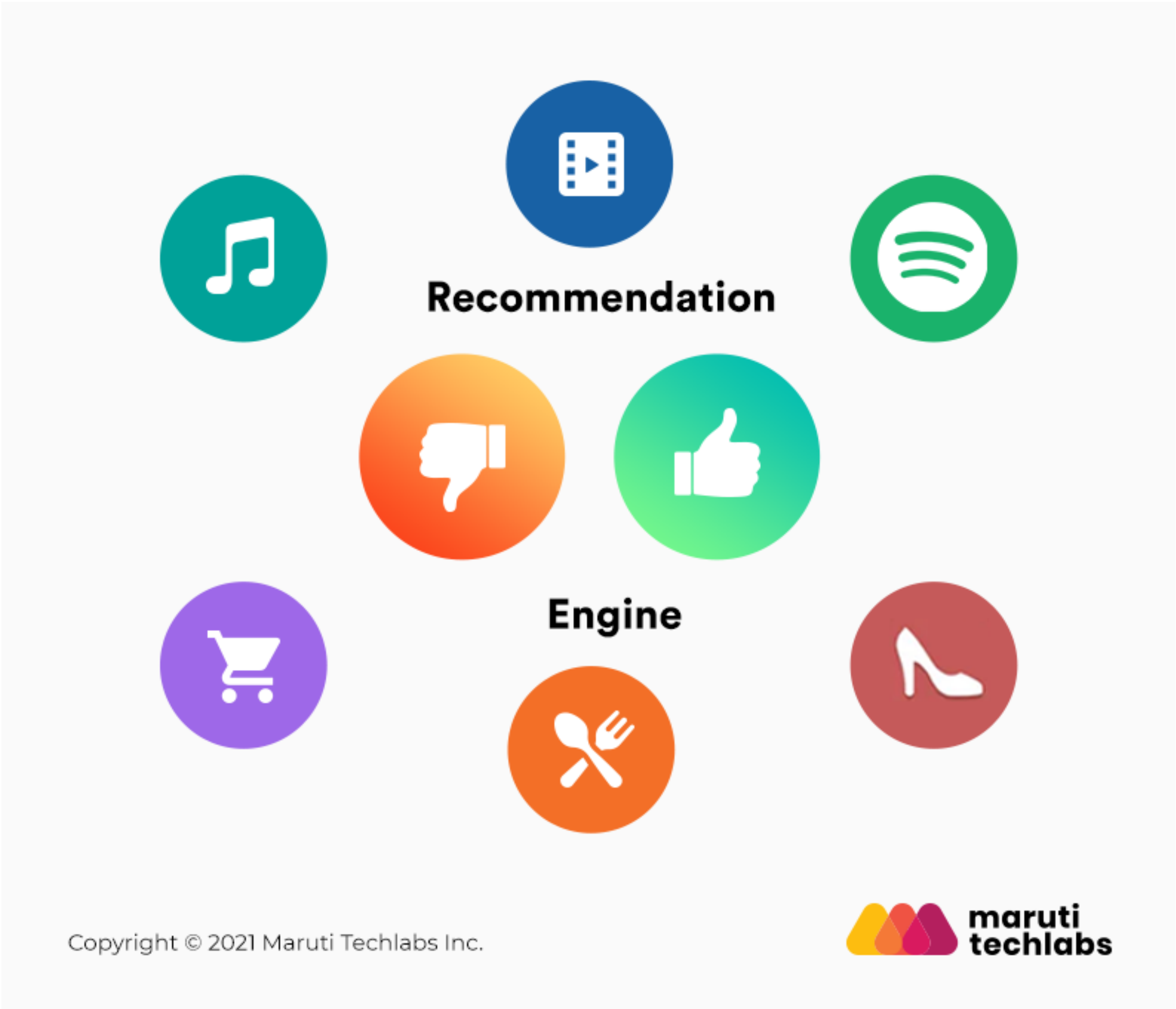
GMM: Gaussian Mixture Model



1. Specify the number of Gaussian components (clusters) “K”
2. Randomly initialize parameters of Gaussian (mean vectors, covariance matrices) and mixing coefficients.
3. Determine the "responsibility" that each Gaussian component has for that point. (responsibility" as a measure of **how likely** the data point came from a particular Gaussian)
4. Update the parameters (mean, covariance, and mixing coefficients) for each Gaussian component
5. Repeat 3 and 4 until converge.

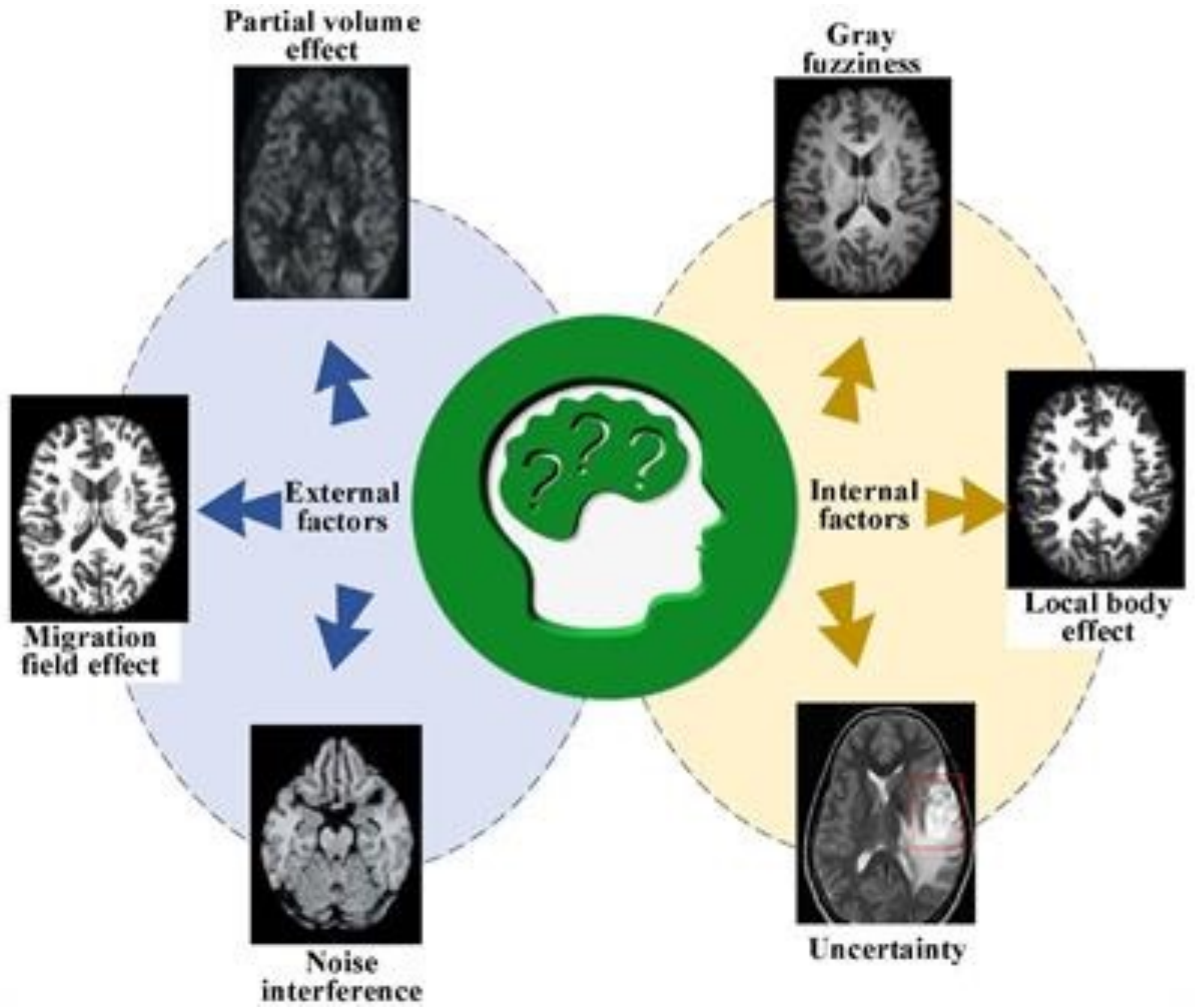
Applications using clustering

Recommendation System



Search Engine

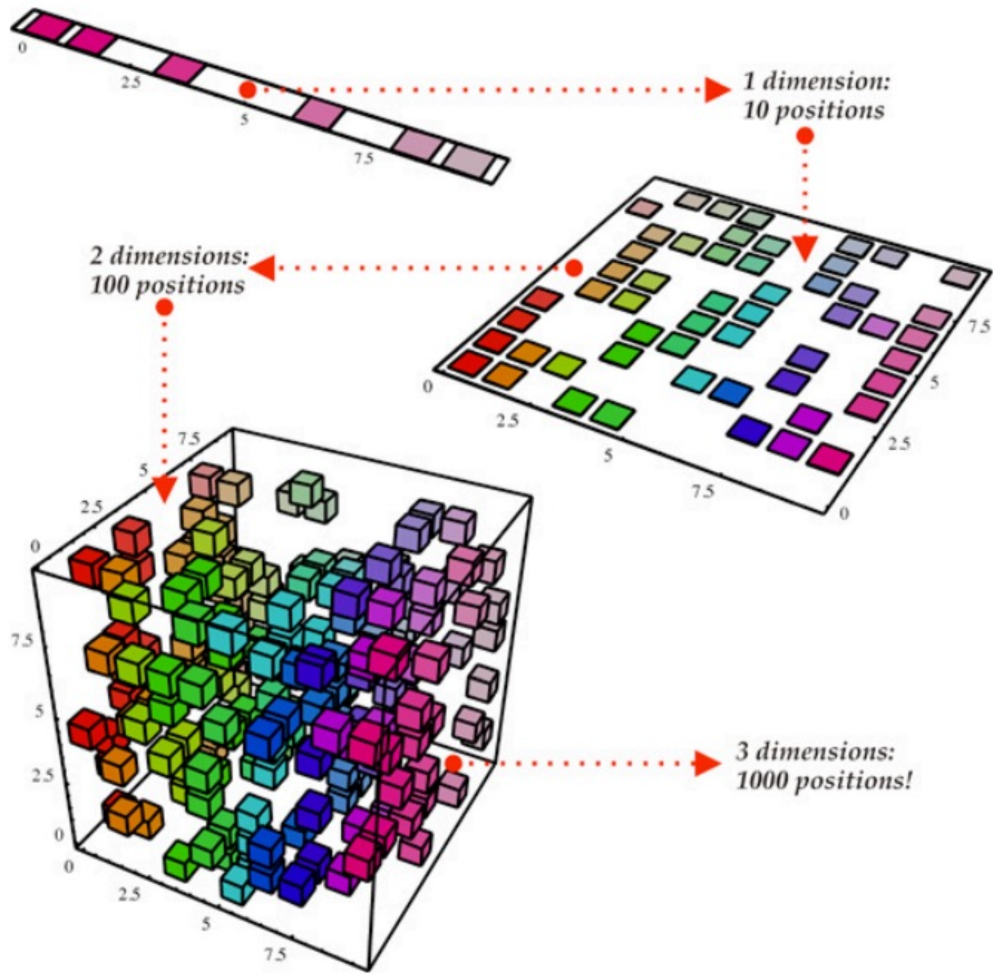
Biomedical Image Analysis



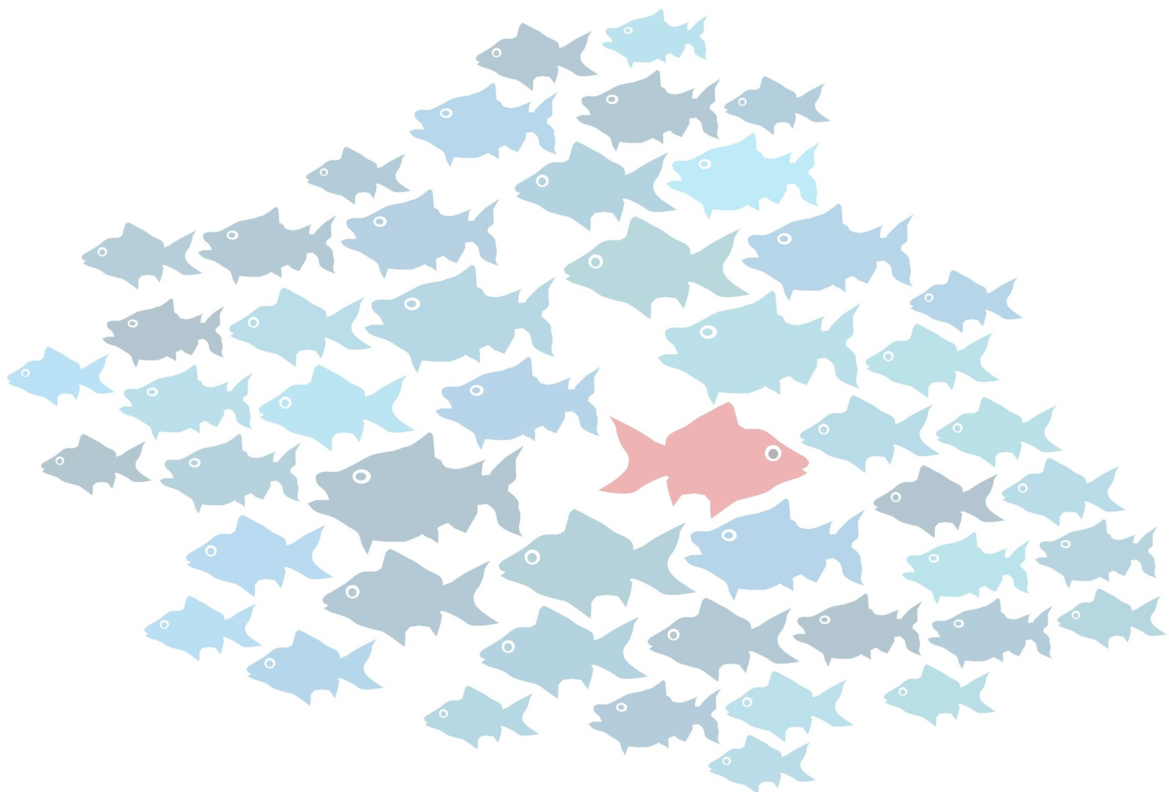
Some common tasks and techniques in unsupervised learning

Clustering

Dimension Reduction

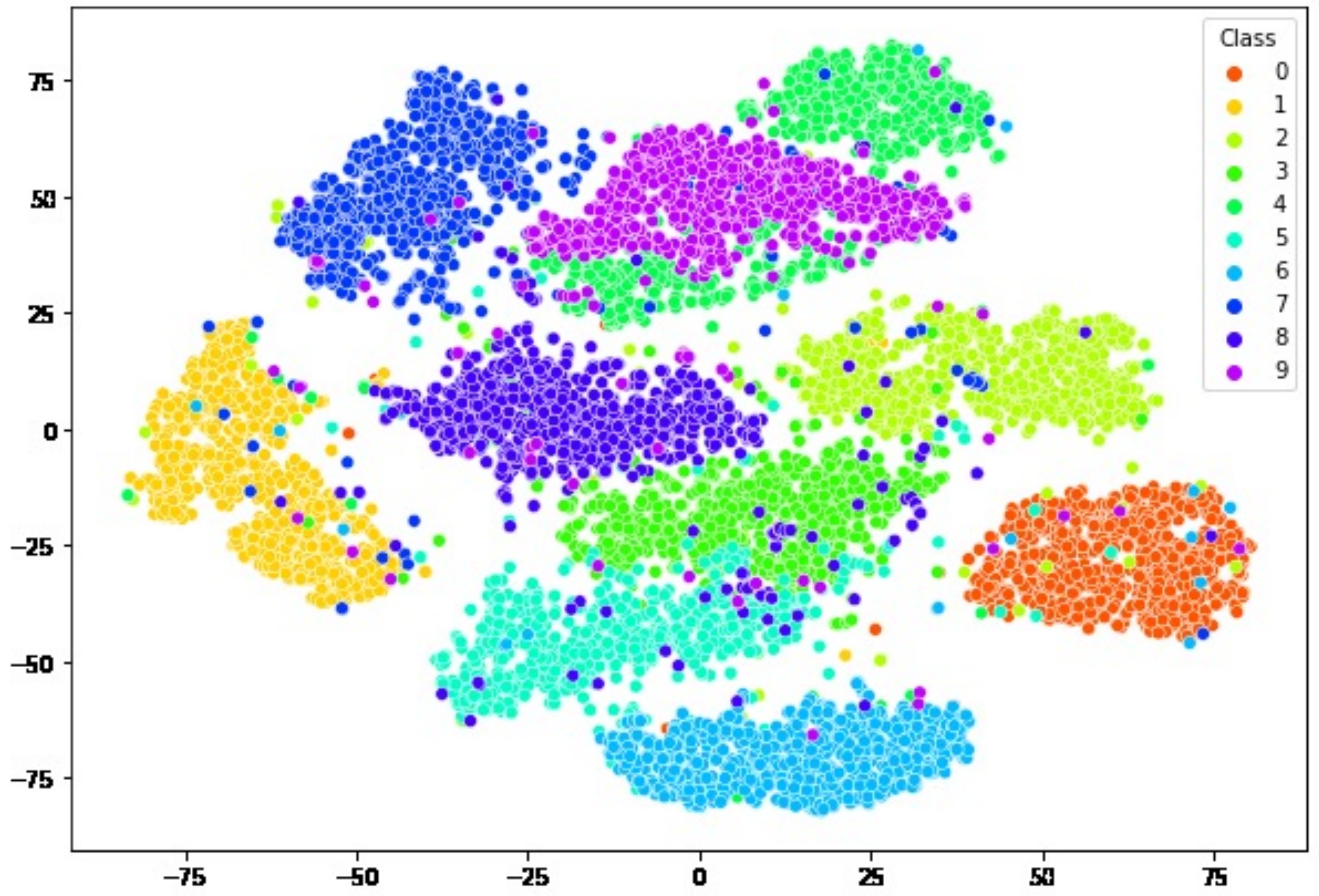


Anomaly Detection



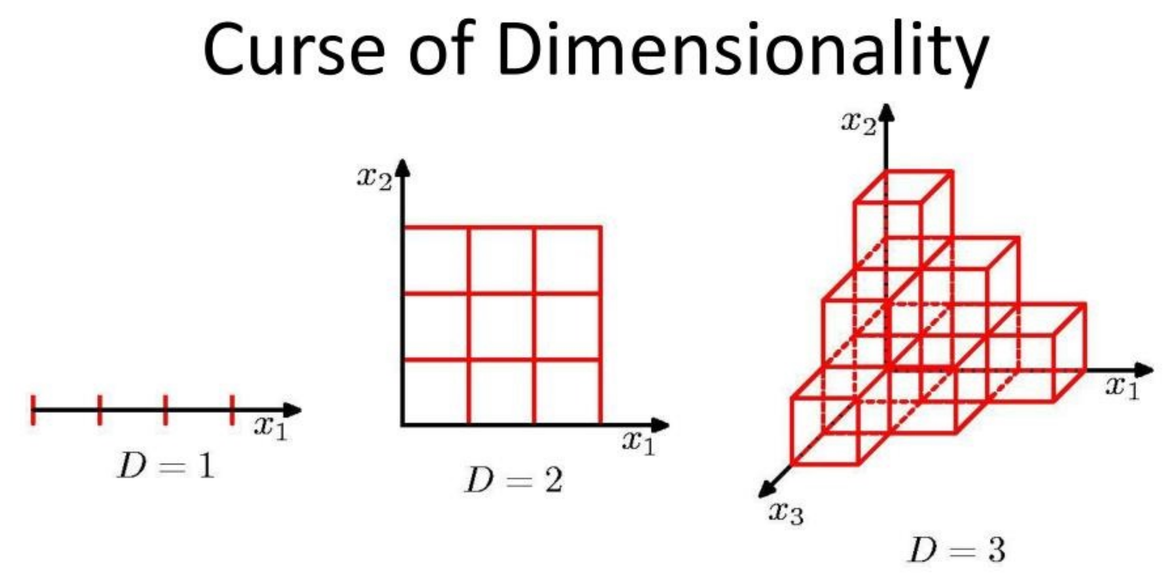
Dimension reduction is a technique used to reduce the number of input features (or dimensions) of a dataset.

Visualization



Improving computational efficiency

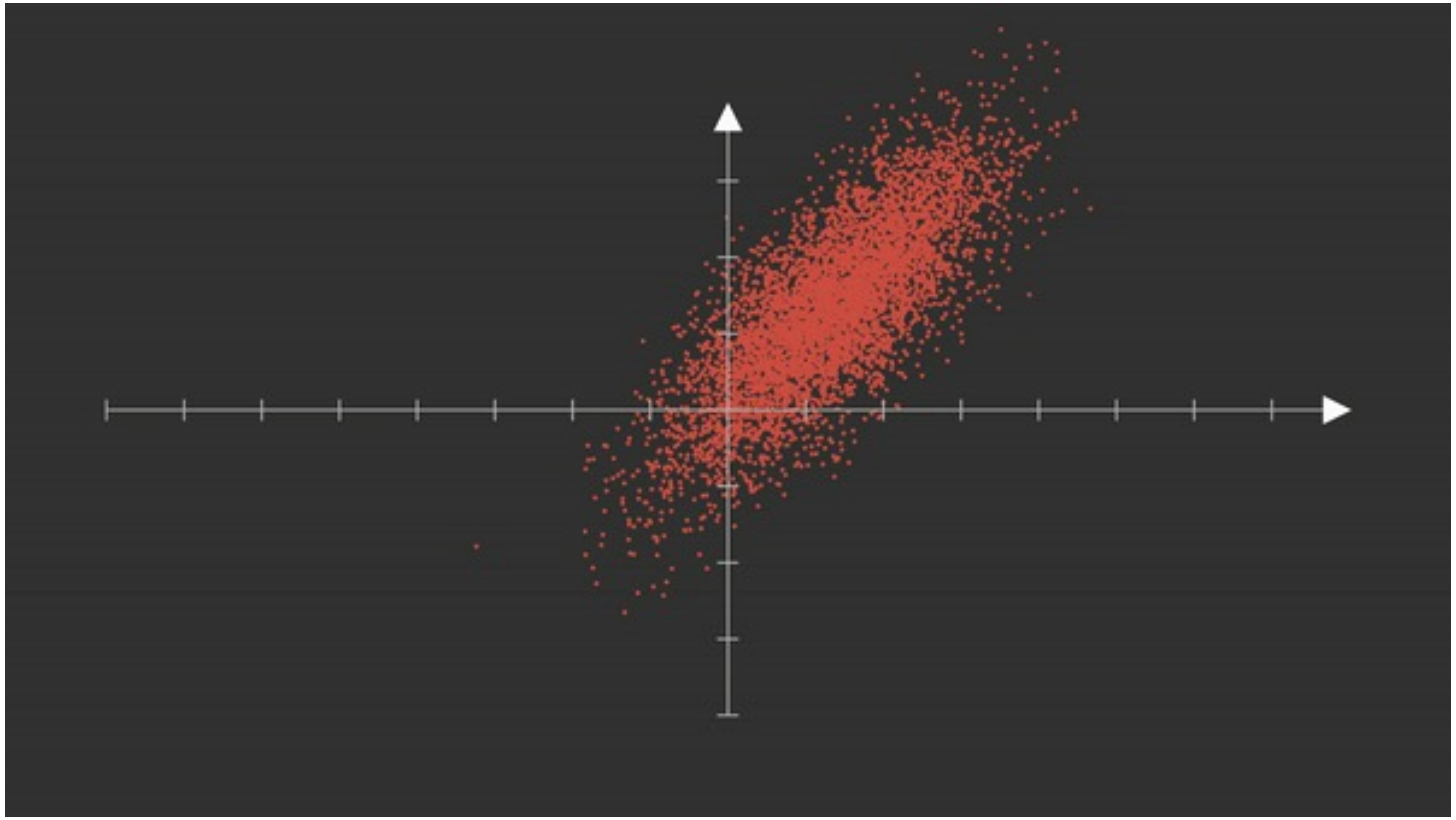
Mitigating the curse of dimensionality



- No. of cells grow exponentially with D
- Need exponentially large no. of training data points
- Not a good approach for more than a few dimensions!

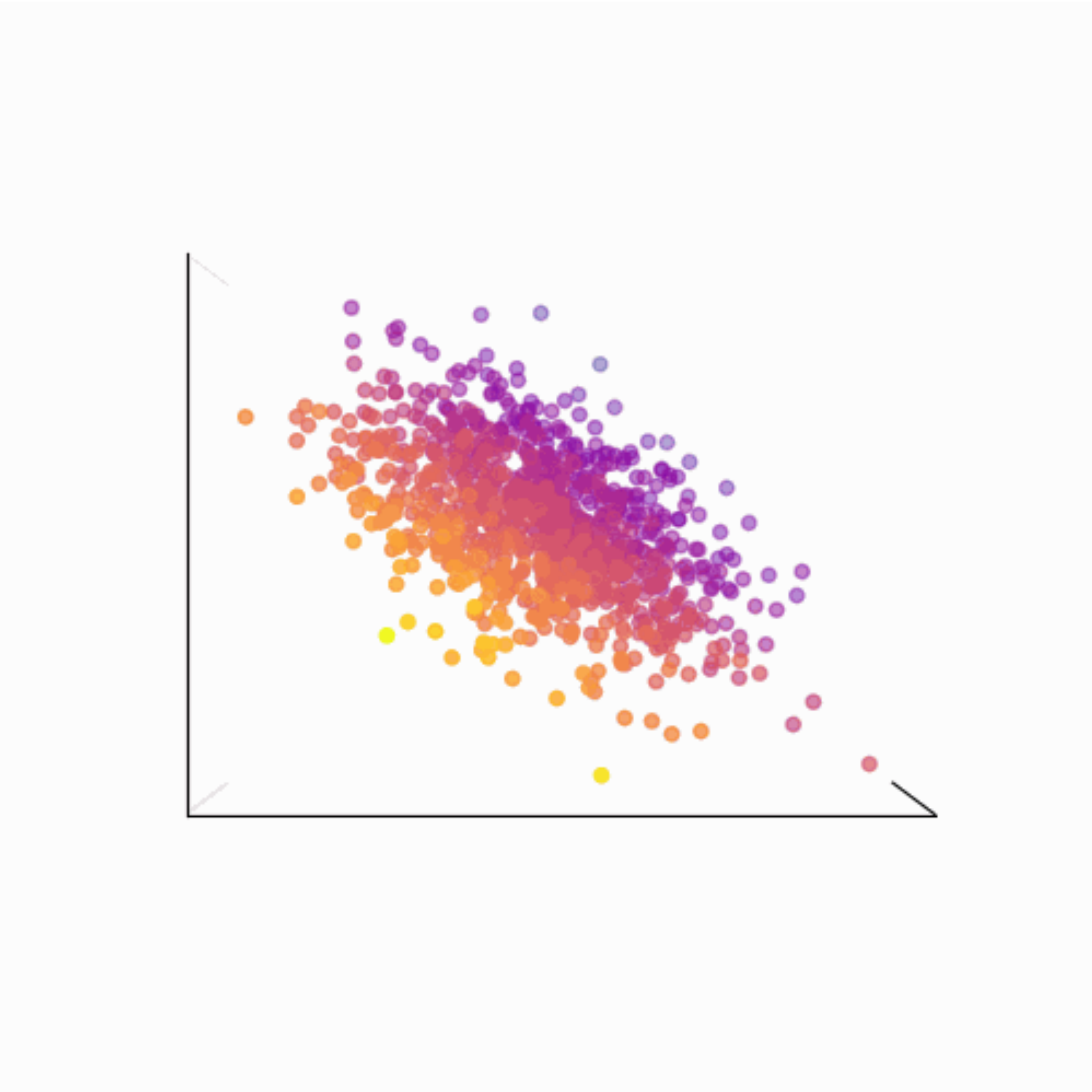
Reference: Christopher M Bishop: Pattern Recognition & Machine Leaning, 2006 Springer

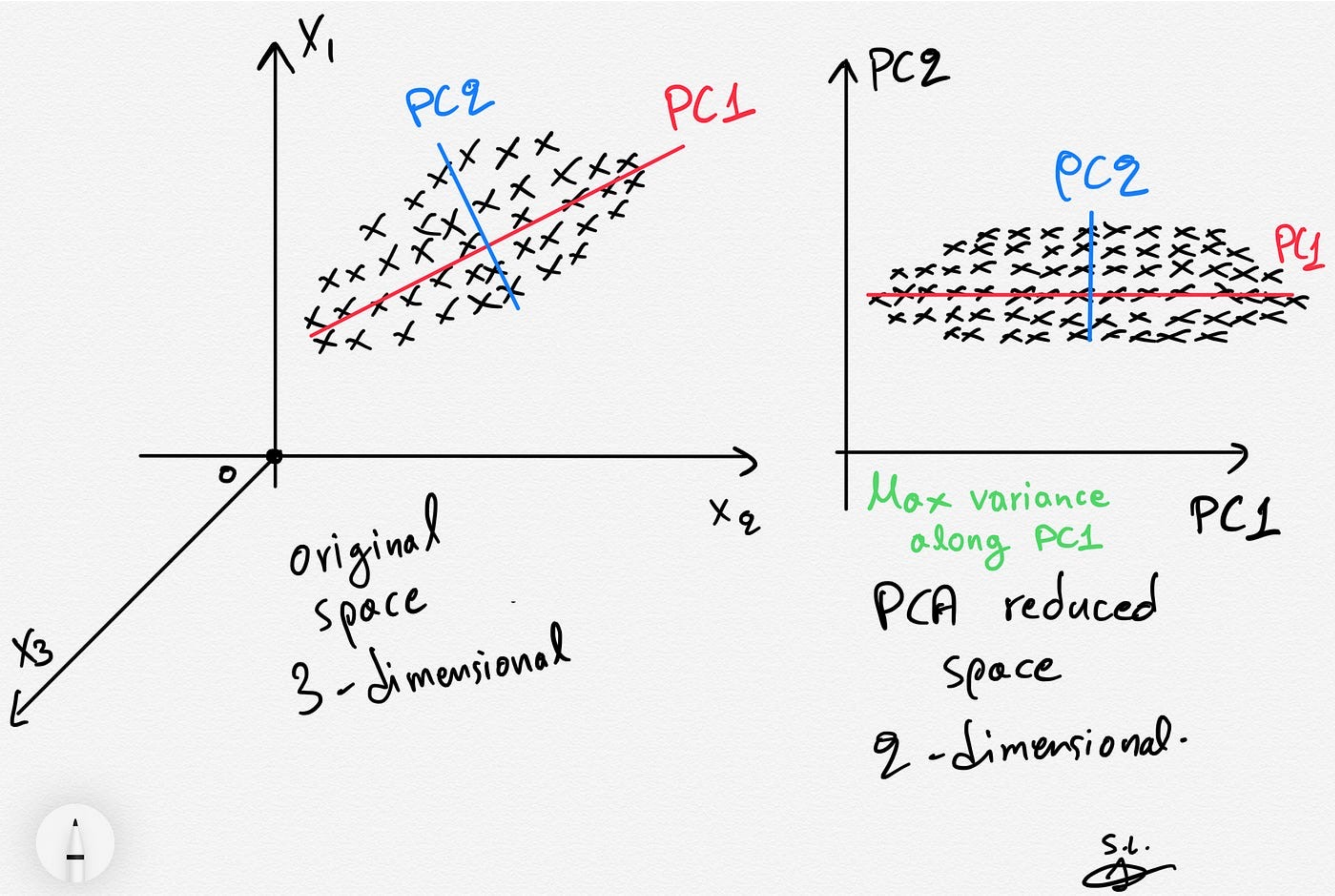
PCA (Principal Component Analysis) is a **projection** while retaining as much of the data's **variance** as possible



https://towardsdatascience.com/eigenvalues-and-eigenvectors-378e851bf372

PCA (Principal Component Analysis) is a **projection** while retaining as much of the data's **variance** as possible

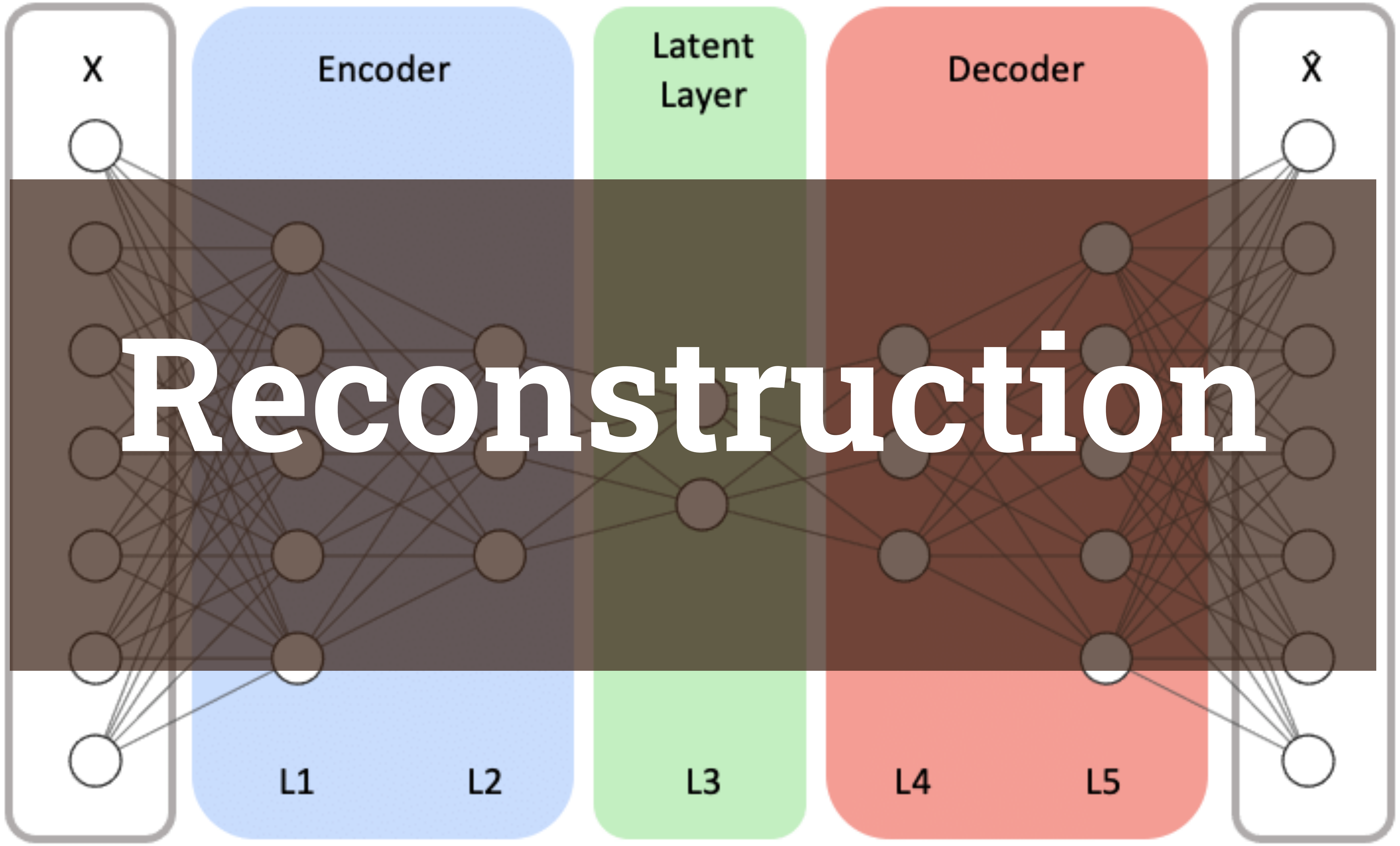




https://towardsdatascience.com/pca-clearly-explained-how-when-why-to-use-it-and-feature-importance-a-guide-in-python-7c274582c37e

Auto-encoder
– A **nonlinear** dimension reduction technique with neural networks

Input data



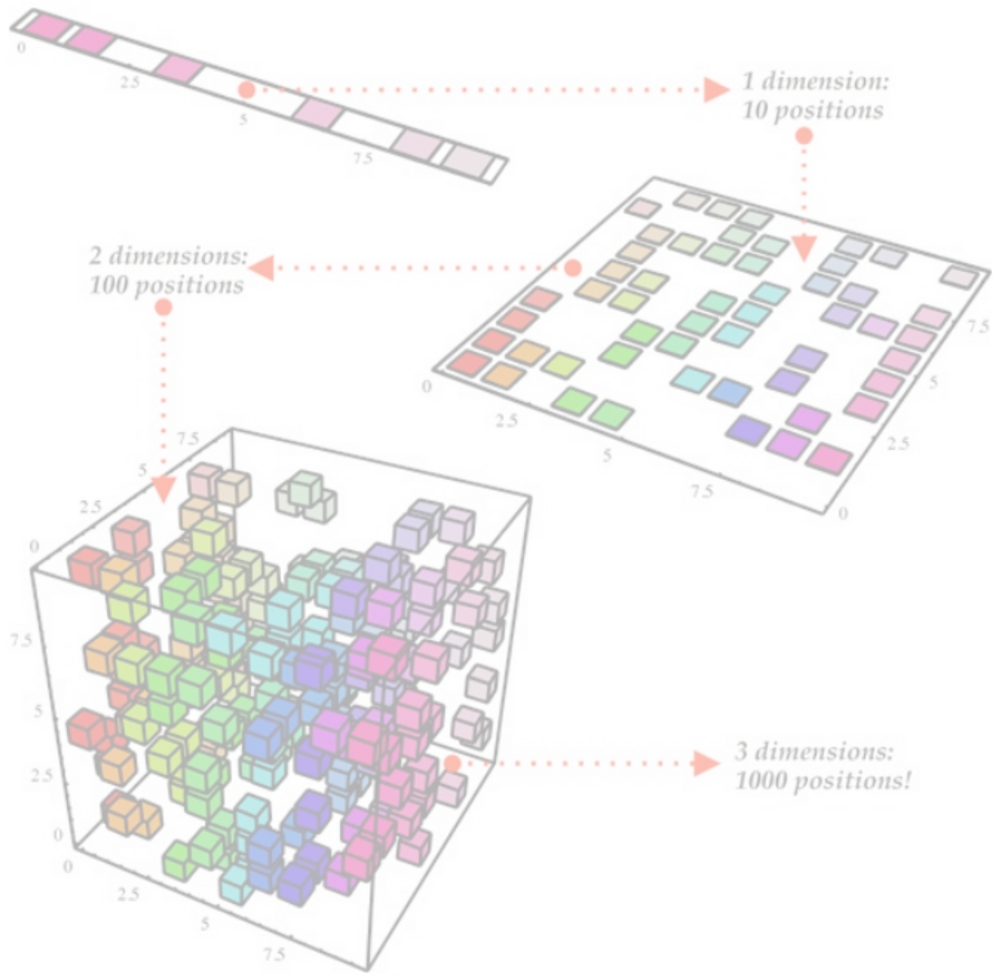
Reconstructed data

https://www.researchgate.net/figure/An-example-autoencoder-model-architecture-with-symmetrical-encoder-and-decoder-networks_fig2_352703131

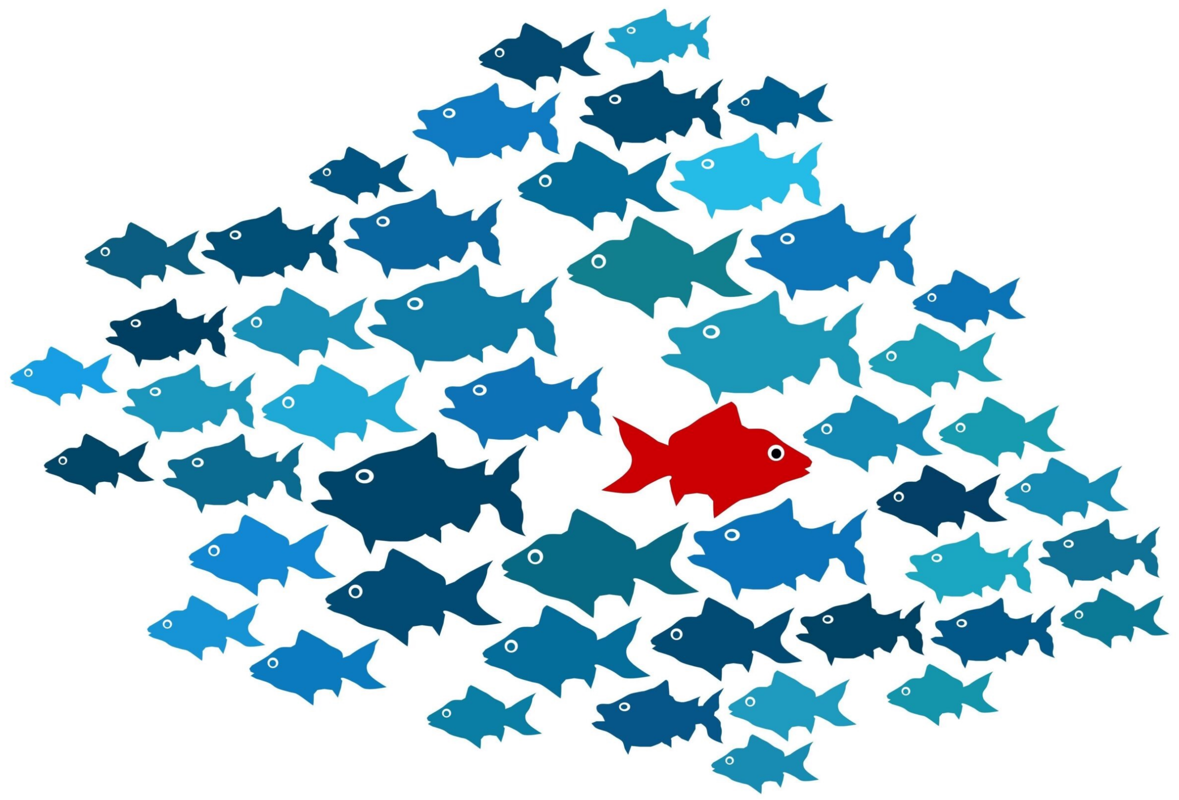
Some common tasks and techniques in unsupervised learning

Clustering

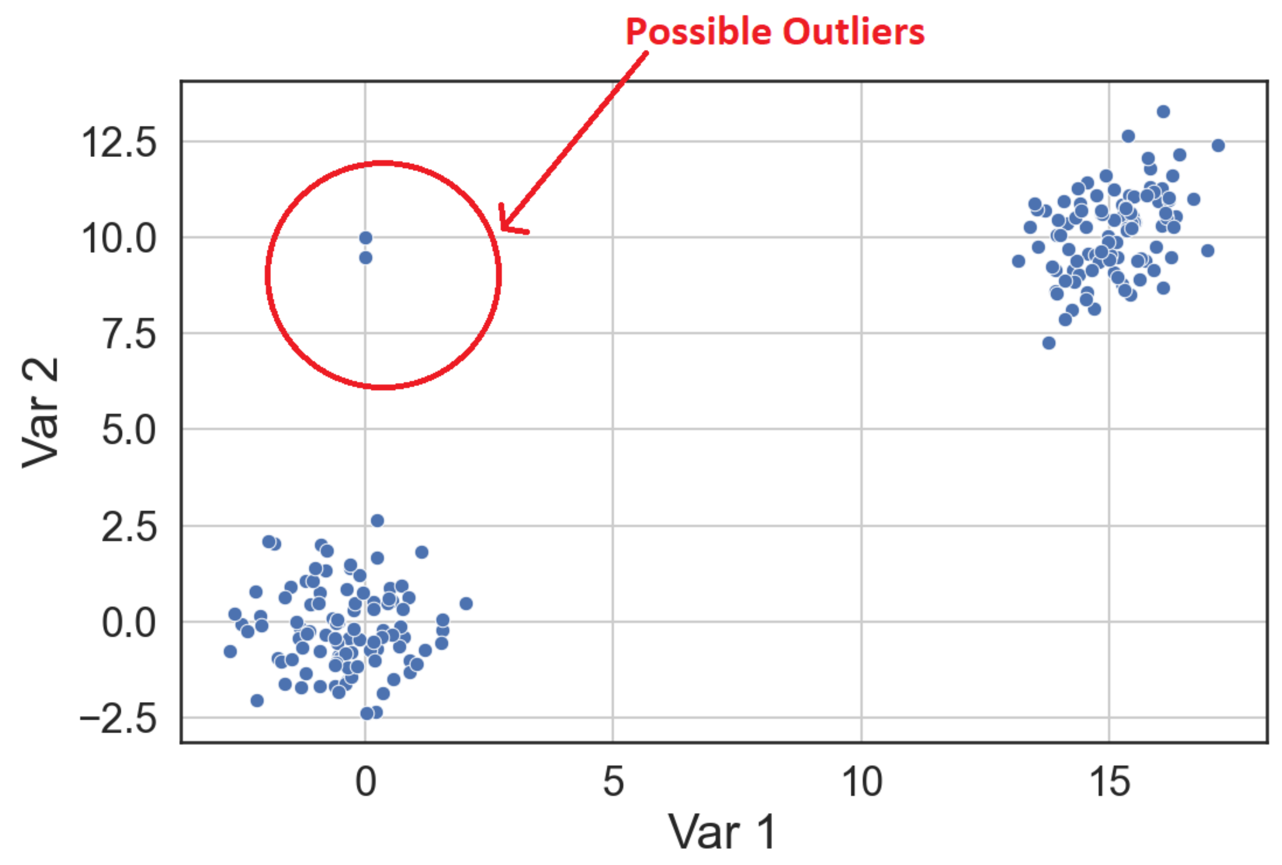
Dimension Reduction

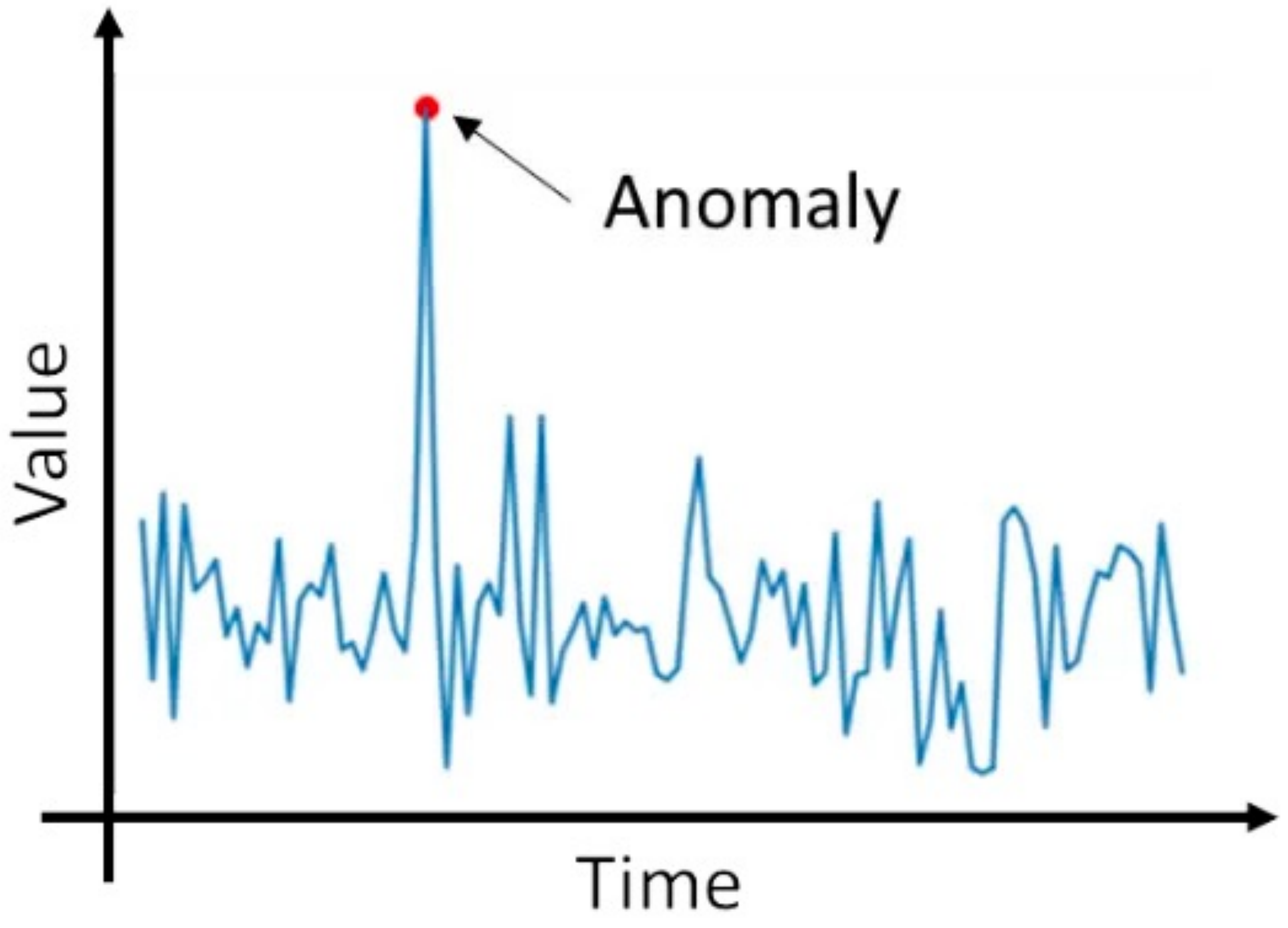


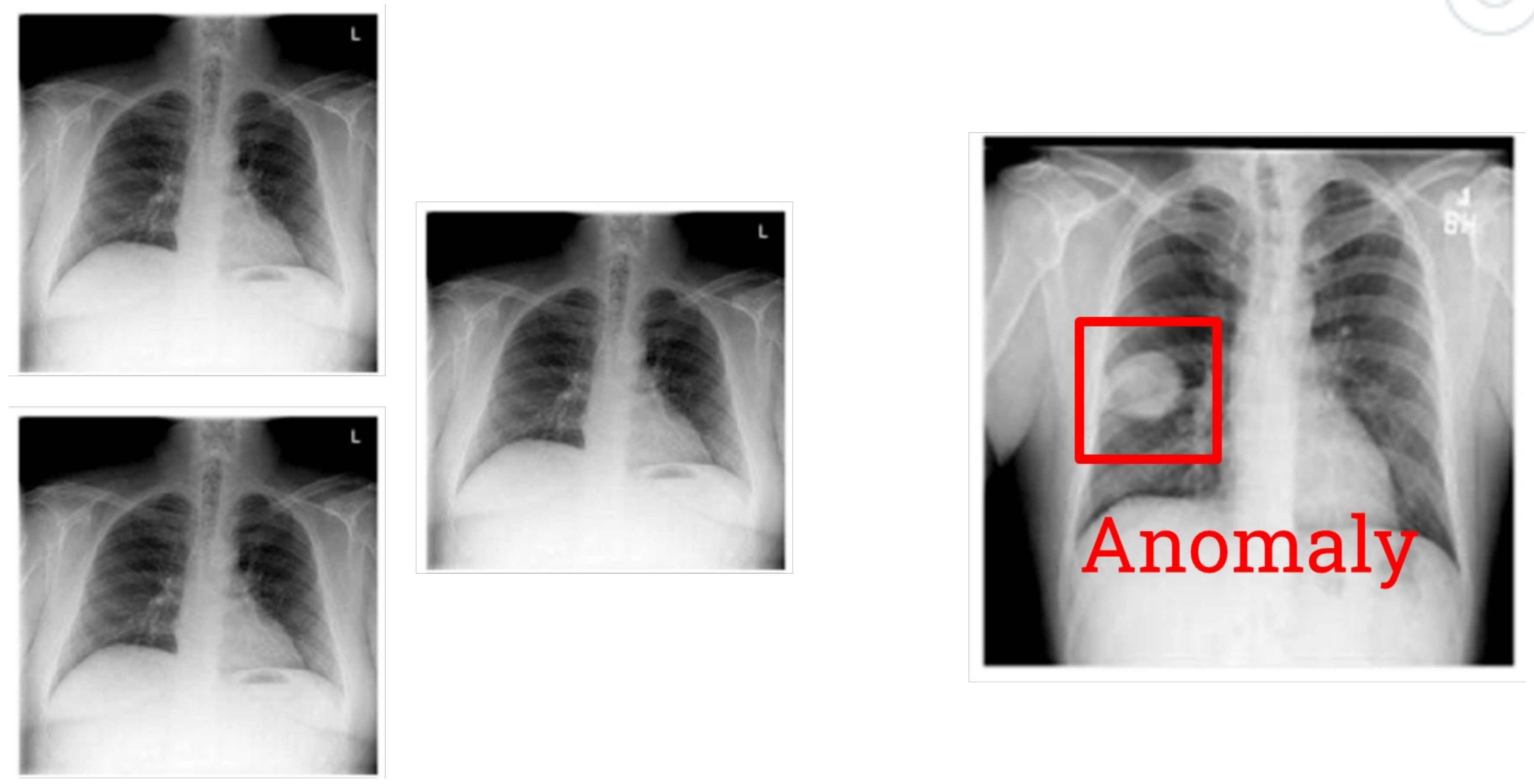
Anomaly Detection



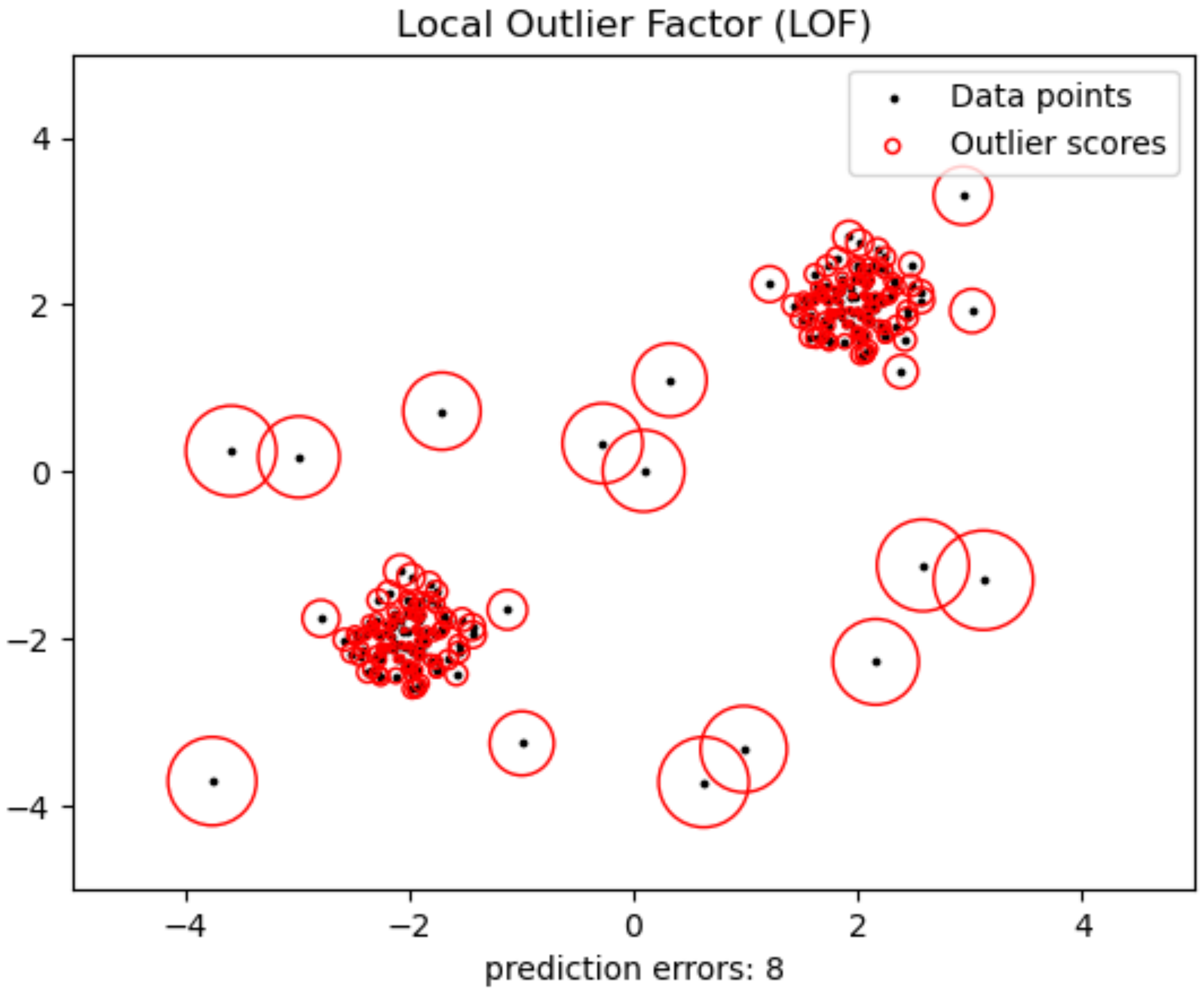
Anomaly detection aims to identify data points or patterns that **deviate significantly from expected behavior**.







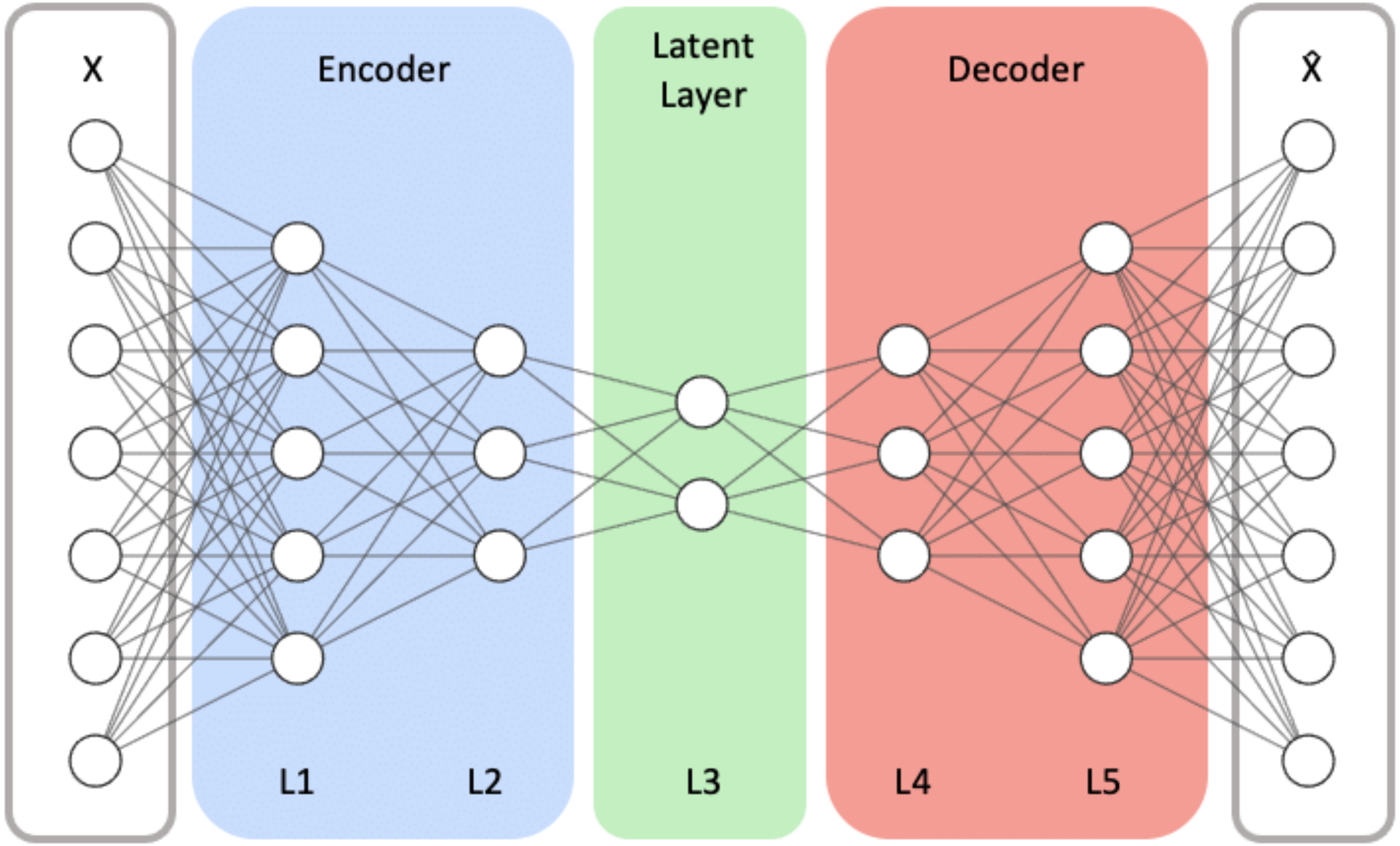
LOF (Local Outlier Factor) detects abnormalities by comparing the **density deviation** of a data point to its neighbors.



1. Specify the “k”, kth nearest neighbor for reachability distance computation.
2. Determine the reachability distance of each point.
3. Compute the local reachability density (LRD), which is the inverse of the average reachability distance of the point from its k neighbors.
4. LOF calculation: For each point, compute the ratio of the average LRD of its neighbors to its own LRD.

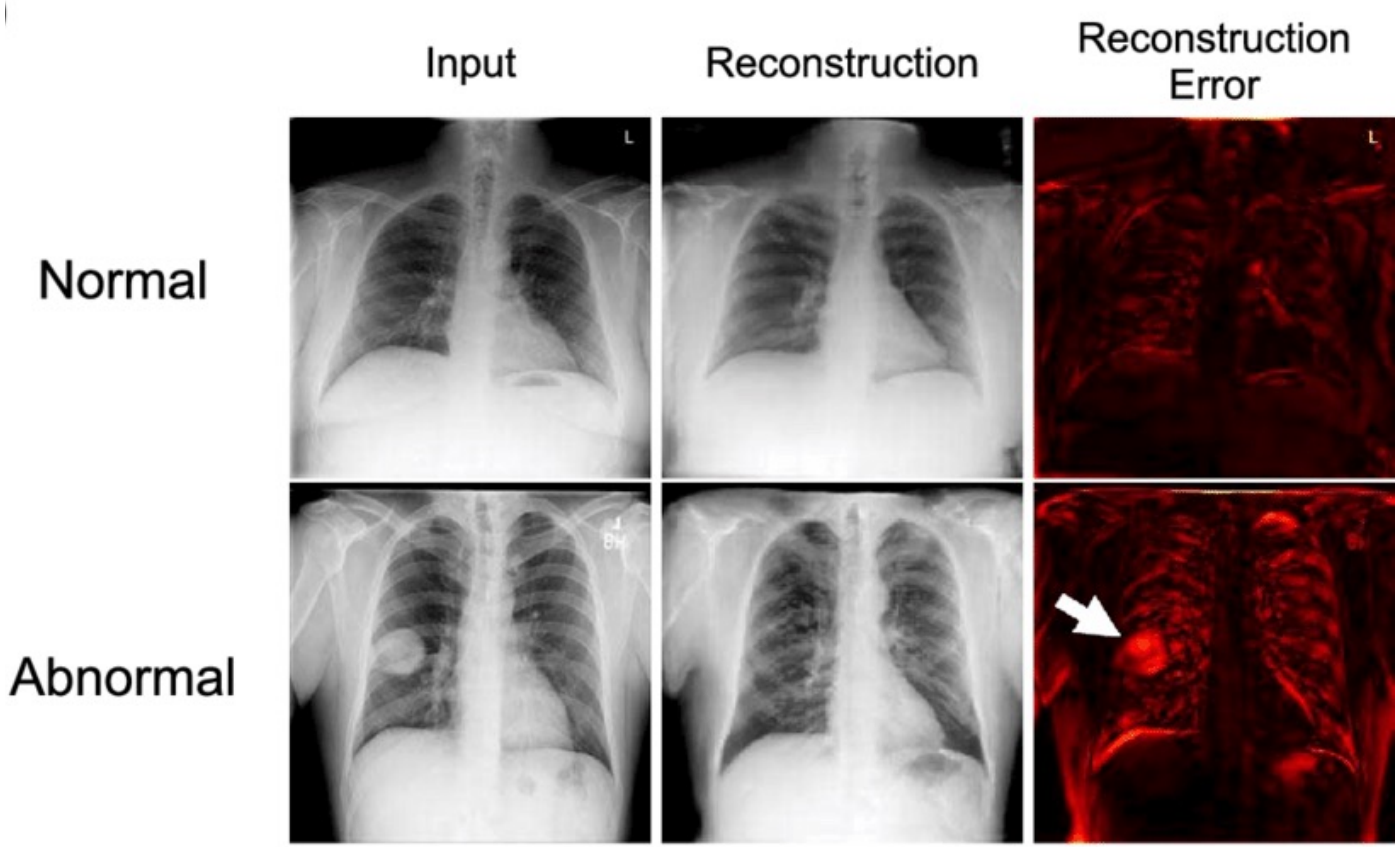
Anomaly detection by using **auto-encoder**

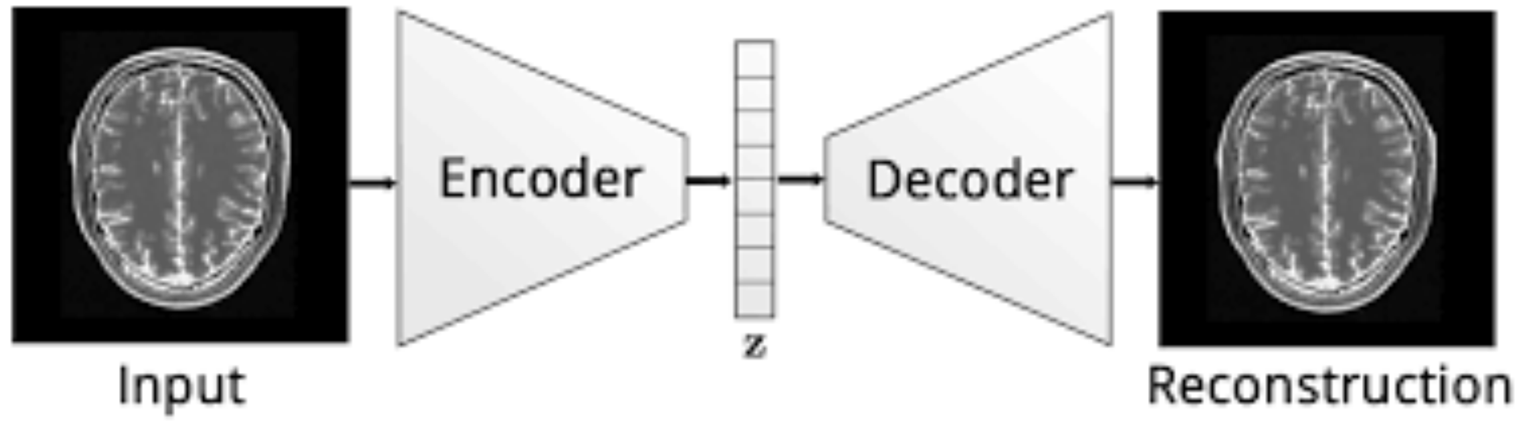
Input data

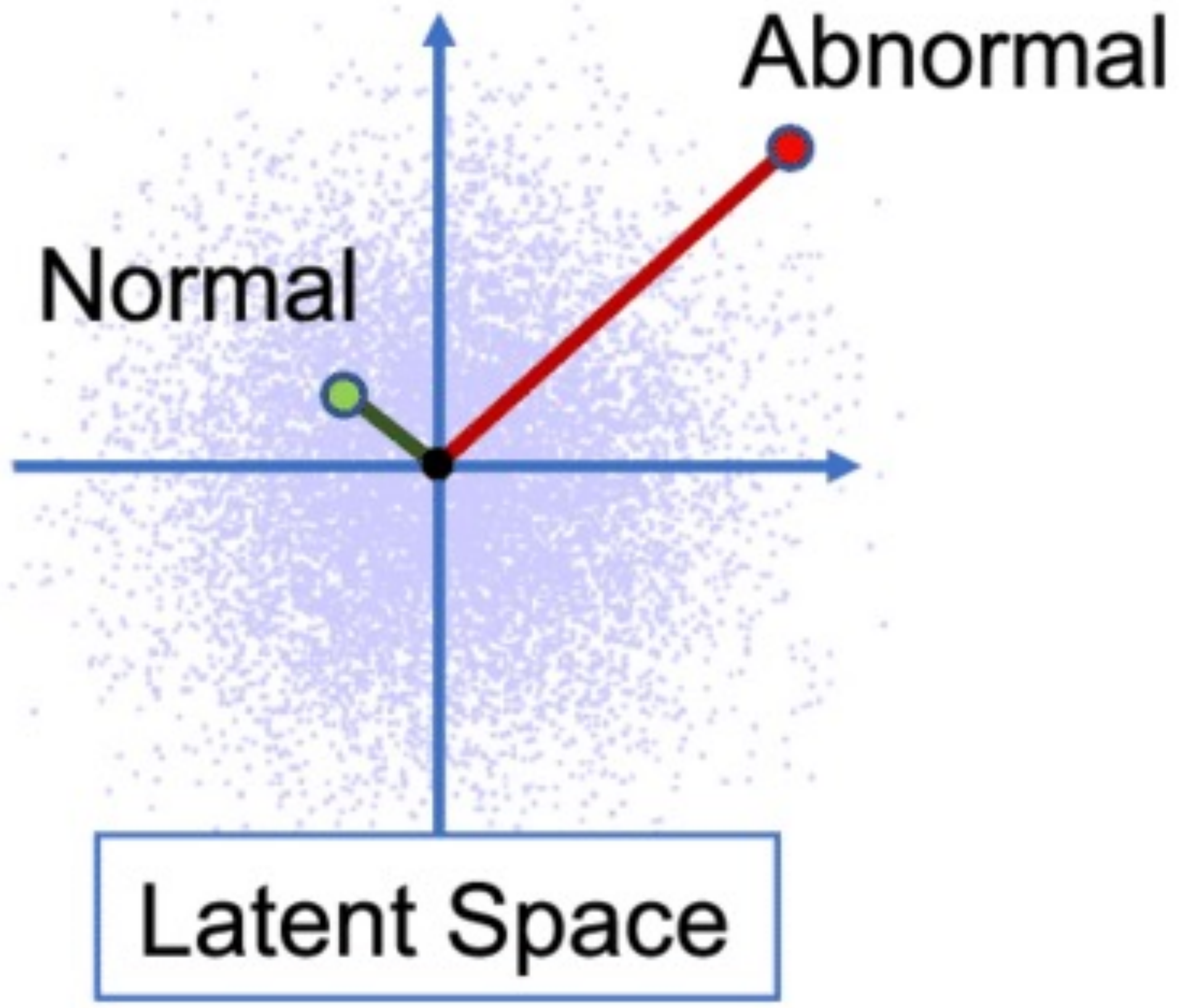


Reconstructed data

Anomaly detection by using **auto-encoder**





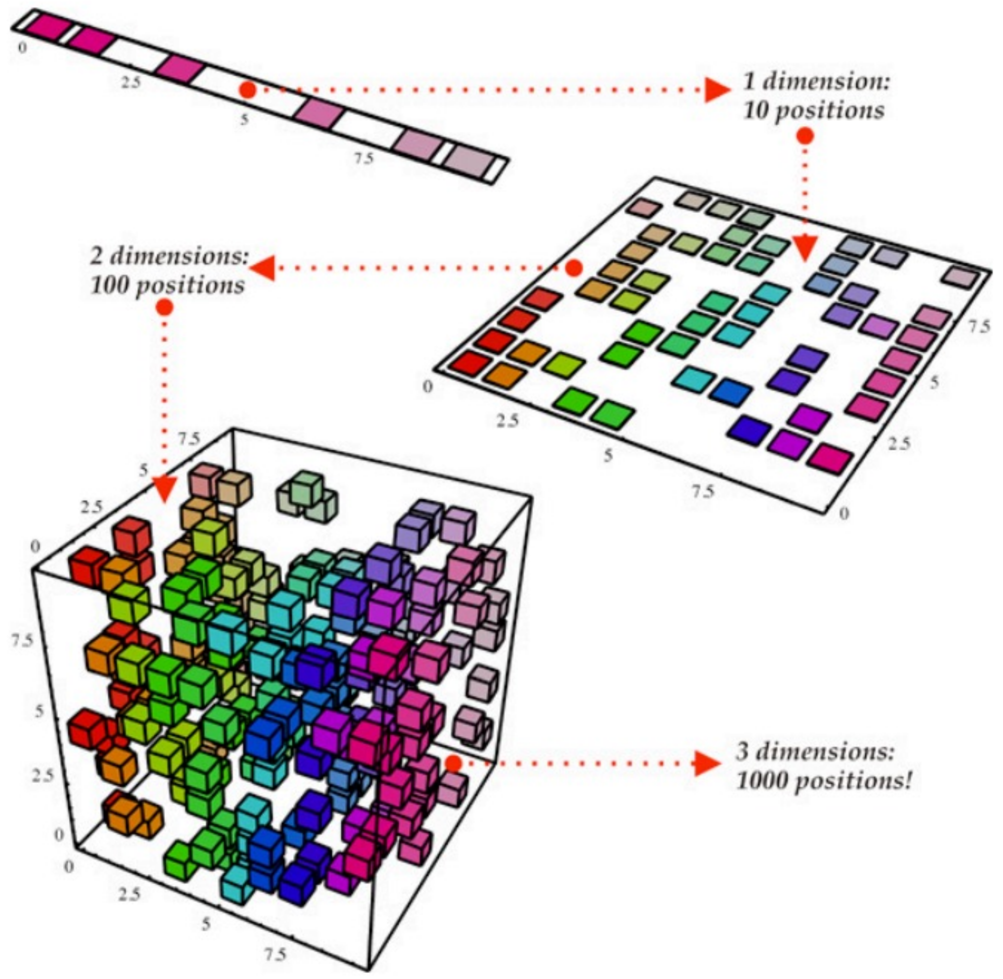


https://www.researchgate.net/figure/Overview-of-our-anomaly-detection-system-a-Anomaly-detection-based-on-reconstruction_fig1_349144159

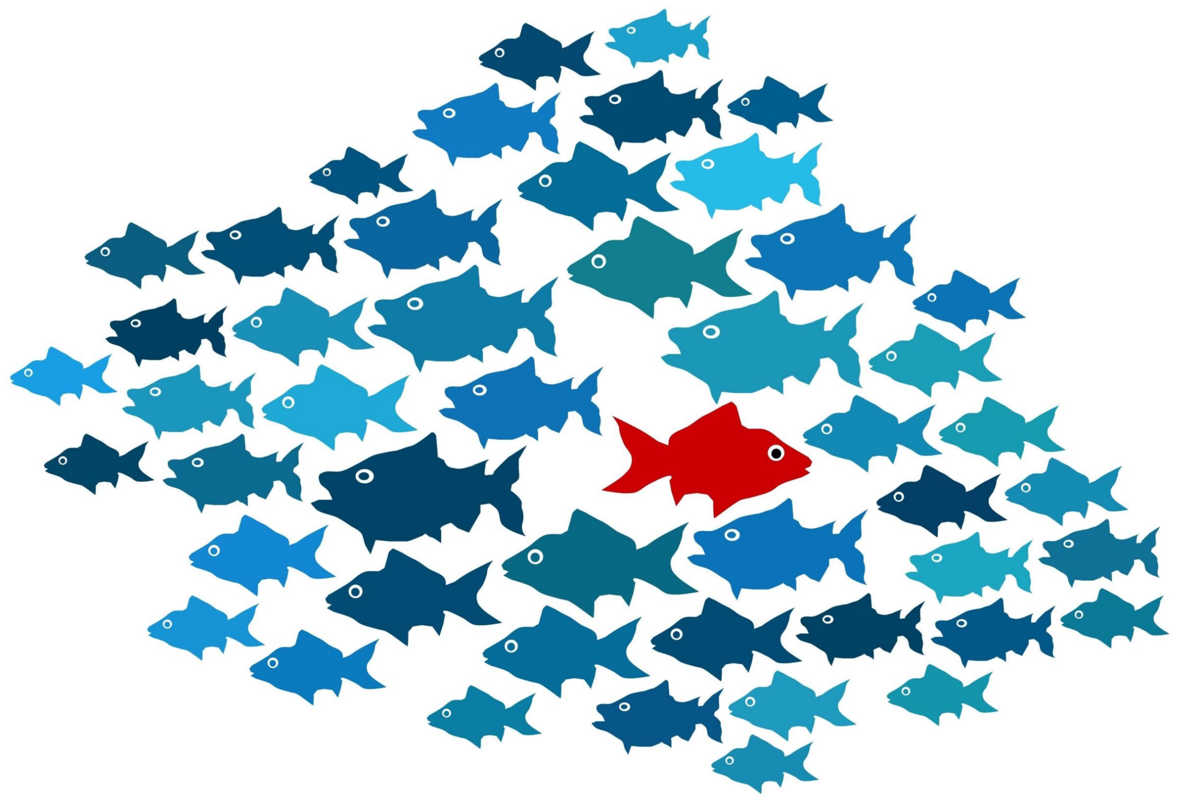
Some common tasks and techniques in unsupervised learning

Clustering

Dimension Reduction



Anomaly Detection



How to **represent** the raw data **effectively** and **accurately**?

Outlines

- What is unsupervised learning
- Methods and applications
- Unsupervised representation (feature) learning

Representation learning is a set of techniques in machine learning where a system **automatically discovers the representations** needed to perform a task (or several tasks).

Representation learning is a set of techniques in machine learning where a system **automatically discovers the representations** needed to perform a task (or several tasks).

Can it be **Unsupervised?**

FACEBOOK AI

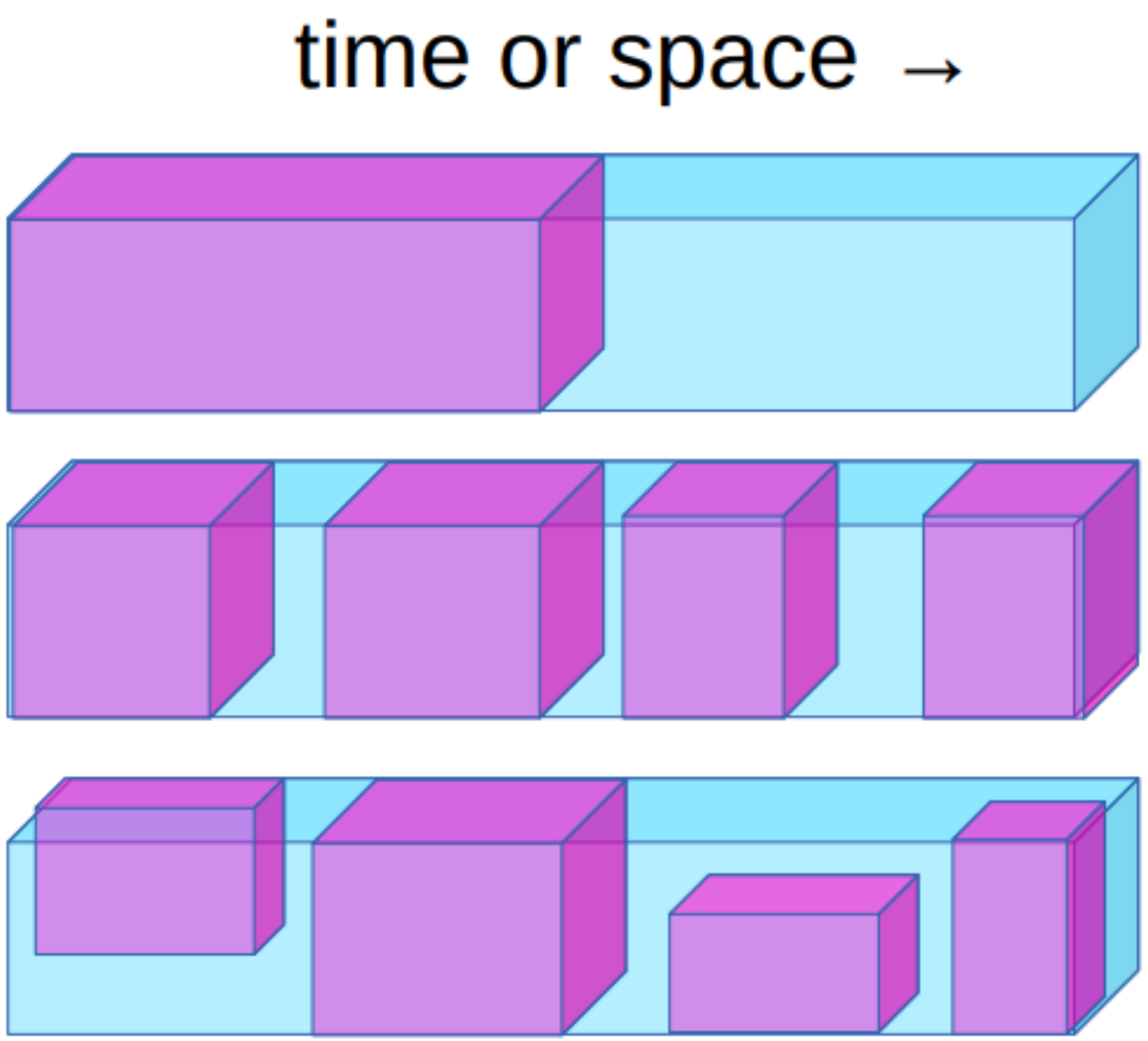
Self-Supervised Learning

Yann LeCun

Predict everything from everything else

Self-Supervised Learning = Filling in the Blanks

- Predict any part of the input from any other part.
- Predict the future from the past.
- Predict the masked from the visible.
- Predict the any occluded part from all available parts.



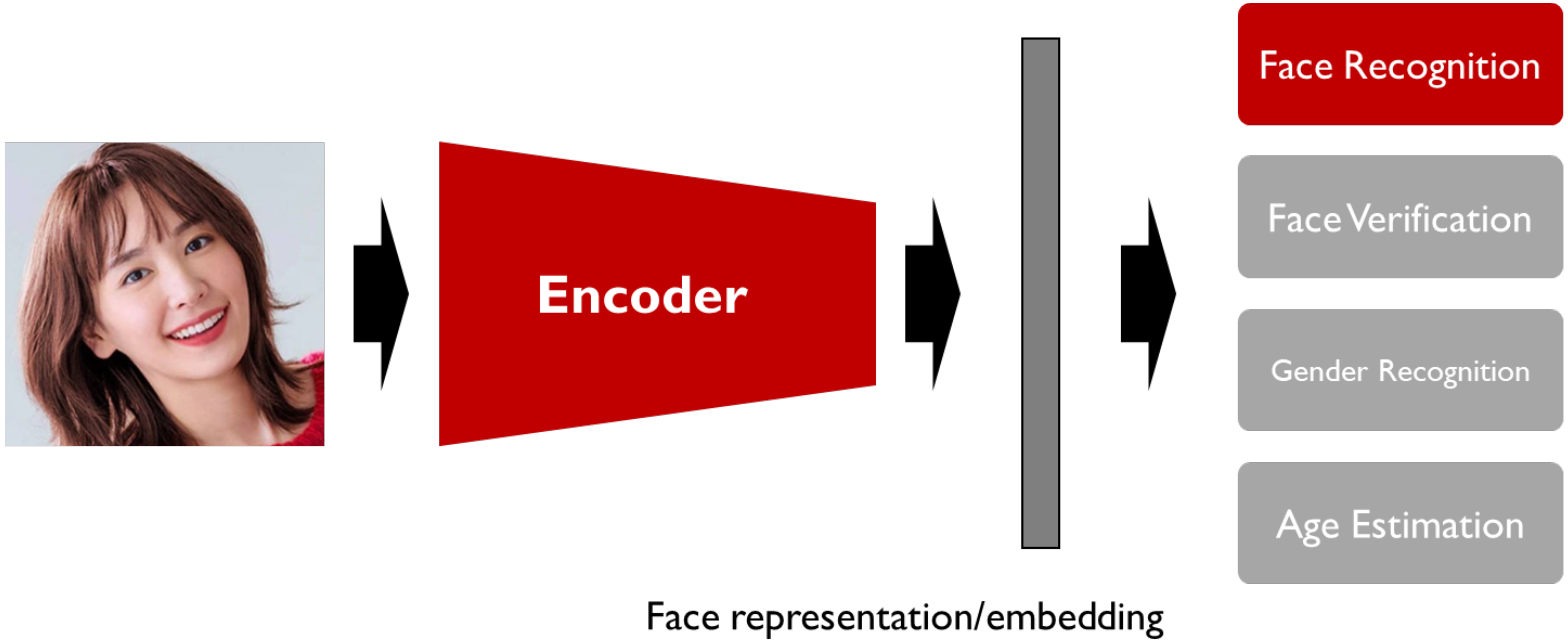
- Pretend there is a part of the input you don't know and predict that.
- Reconstruction = SSL when any part could be known or unknown

How Much Information is the Machine Given during Learning?

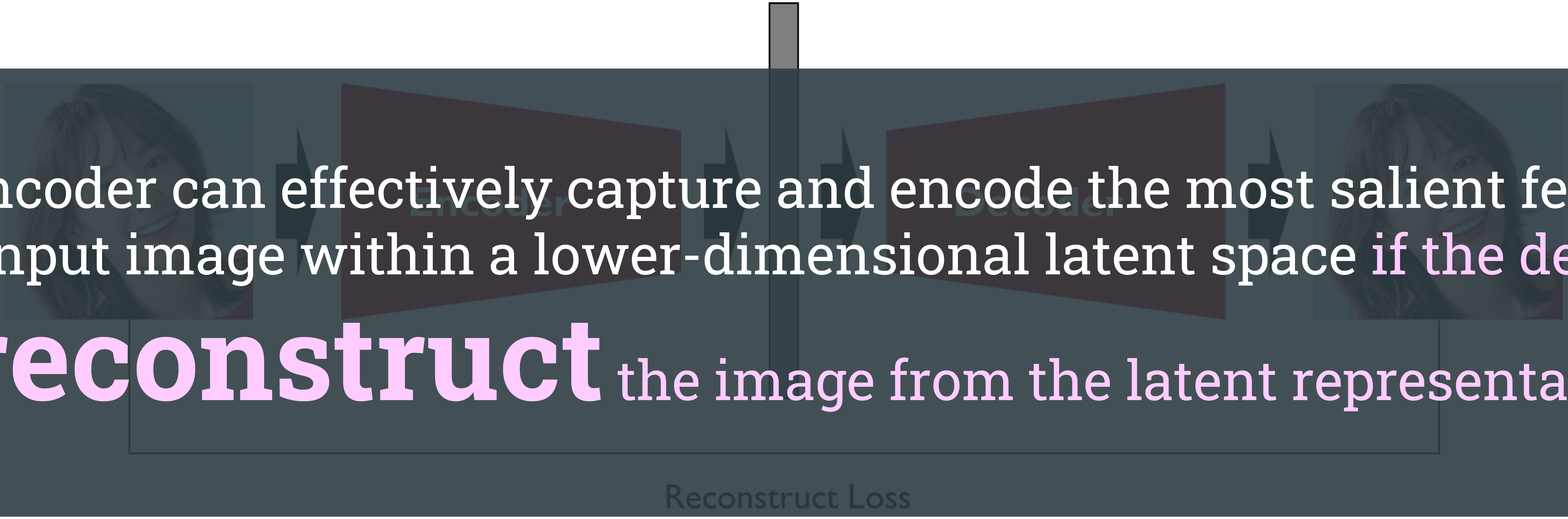
- **“Pure” Reinforcement Learning (cherry)**
  - The machine predicts a scalar reward given once in a while.
  - **A few bits for some samples**
- **Supervised Learning (icing)**
  - The machine predicts a category or a few numbers for each input
  - Predicting human-supplied data
  - **10→10,000 bits per sample**
- **Self-Supervised Learning (cake génoise)**
  - The machine predicts any part of its input for any observed part.
  - Predicts future frames in videos
  - **Millions of bits per sample**

Self-supervised learning is a technique to learn a suitable feature representation for the downstream task by a well-designed pretext/surrogate task without relying on explicit external labels or annotations.

Representation in Face Recognition Applications

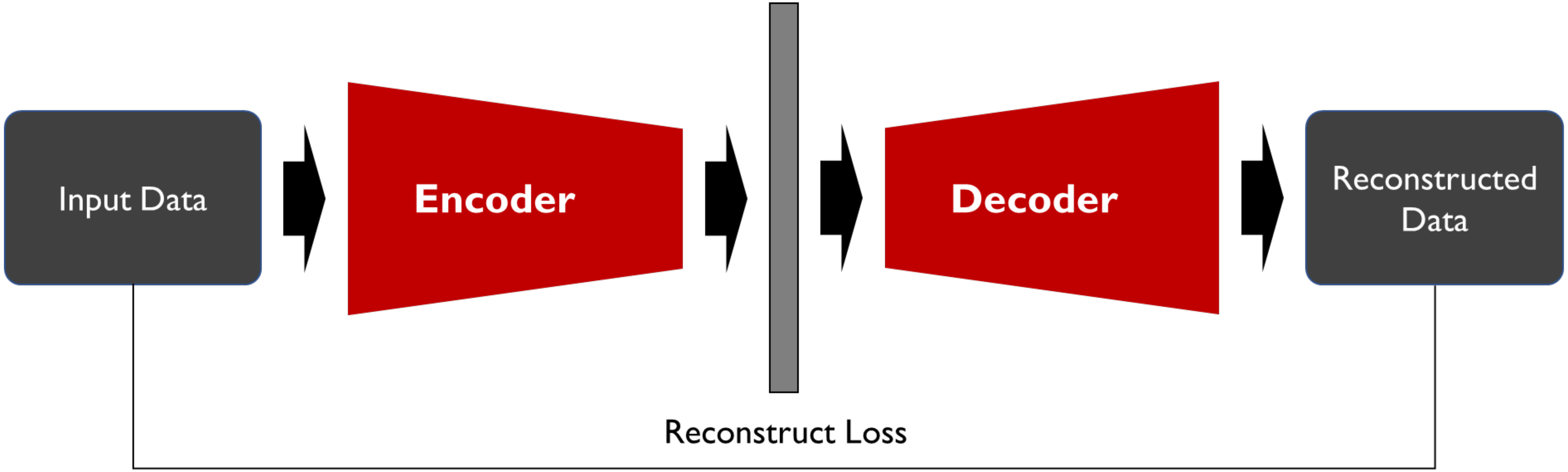


What is the Pretext/Surrogate task?

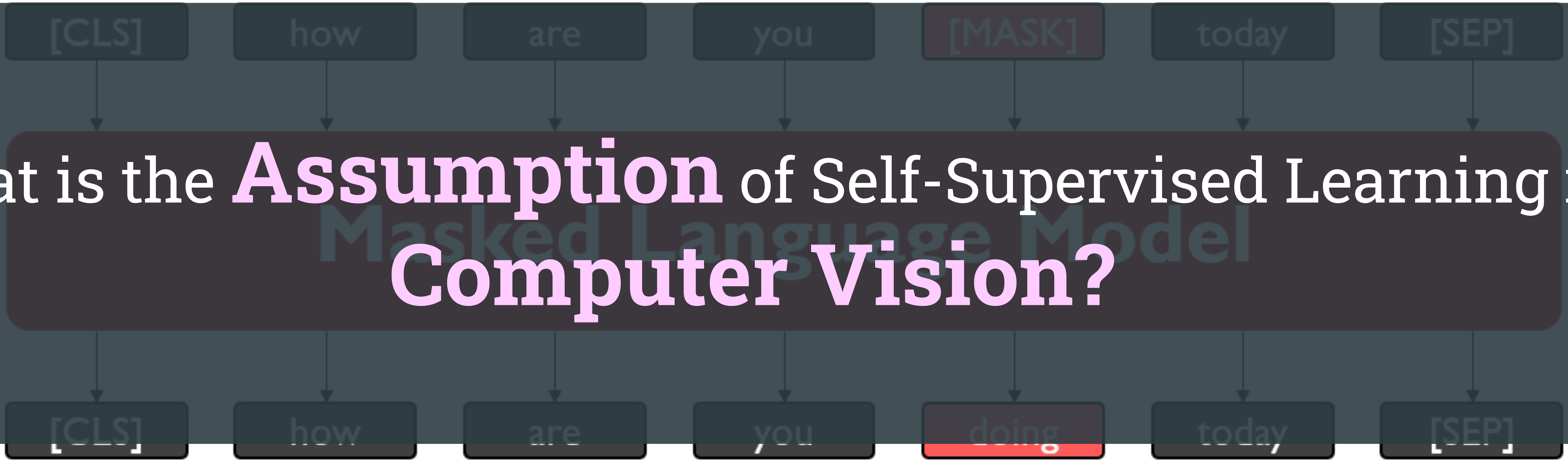


Auto-encoder can effectively capture and encode the most salient features of the input image within a lower-dimensional latent space if the decoder can **reconstruct** the image from the latent representation.

The pretext task in self-supervised learning typically relies on certain **assumptions** related to the downstream application.



The pretext task in self-supervised learning typically relies on certain assumptions related to the downstream application.



What is the **Assumption** of Self-Supervised Learning in **Computer Vision?**

Different types of SSL in Computer Vision

Generative vs Contrastive Methods

Contemporary self-supervised learning methods can roughly be broken down into two classes of methods:

Generative / Predictive

Loss measured in the output space

Examples: Colorization, Auto-Encoders

Contrastive

Loss measured in the representation space

Examples: TCN, CPC, Deep-InfoMax

Source: Contrastive Self-Supervised Learning, https://ankeshanand.com/blog/2020/01/26/contrative-self-supervised-learning.html

Contrastive Learning in Humans

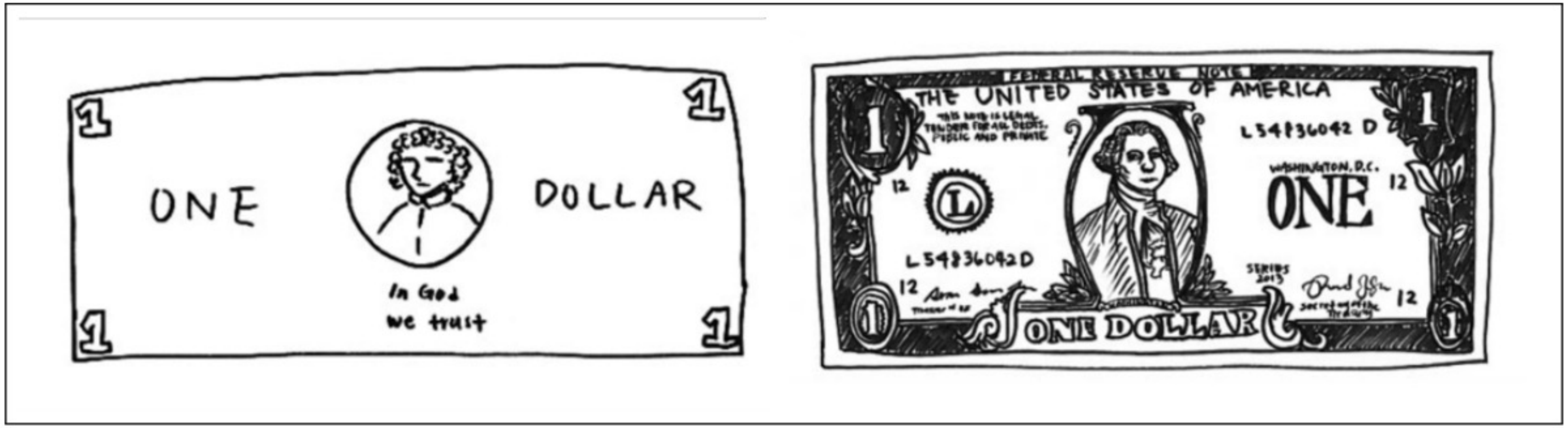


Figure 1: Fig. Left: Drawing of a dollar bill from memory. Right: Drawing subsequently made with a dollar bill present. Image source: Epstein, 2016.

Which one is most similar to the target image?

Target Image

Which one is most similar to the target image?

Target Image

The most similar to an image should be its
Augmentations(different views)!

Positive Pair | Negative Pair

$similarity(f($  $), f($  $)) \gg similarity(f($  $), f($  $))$

Positive Pair | Negative Pair

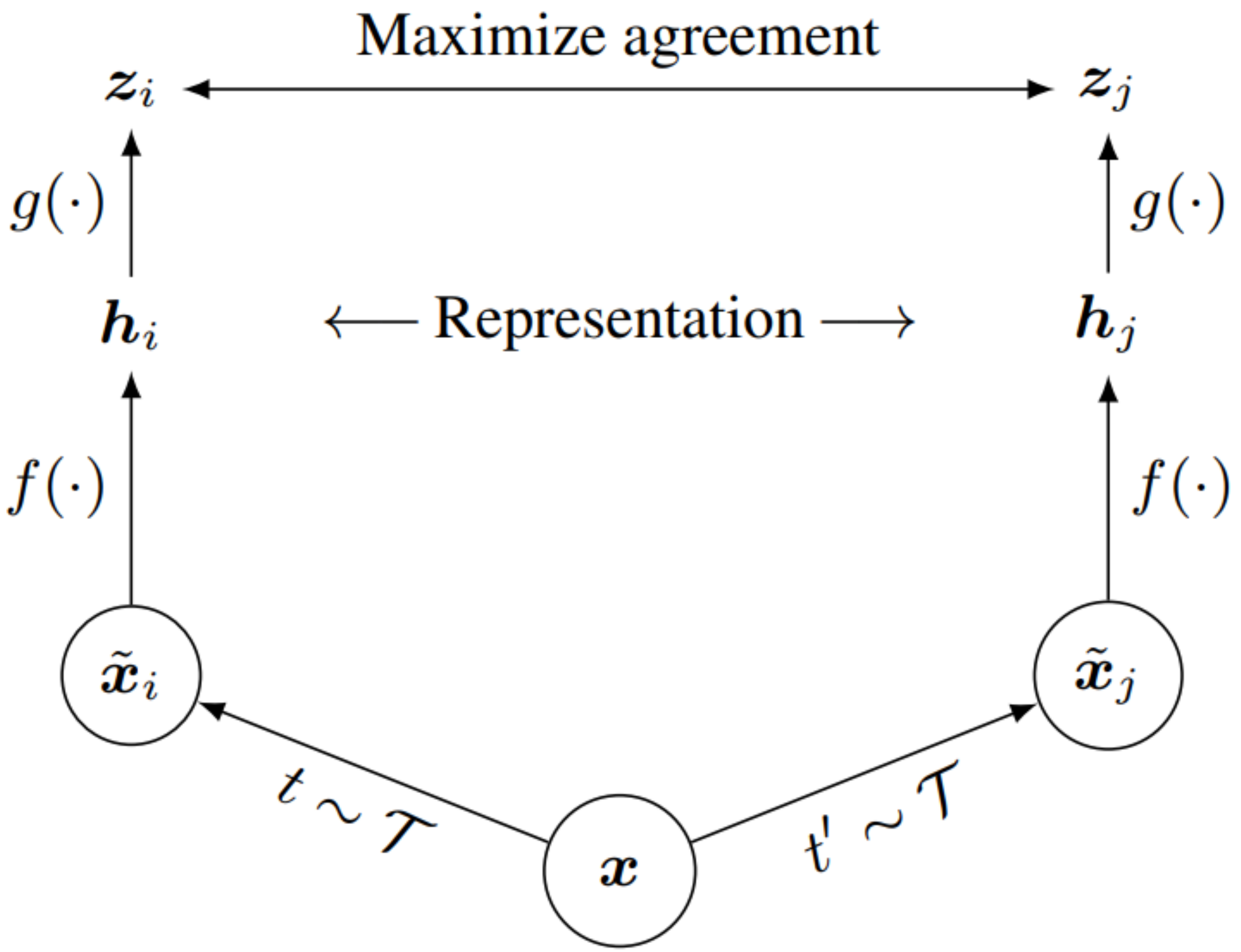
$similarity(f($  $), f($  $)) \gg similarity(f($  $), f($  $))$

SimCLR

NT-Xent Loss

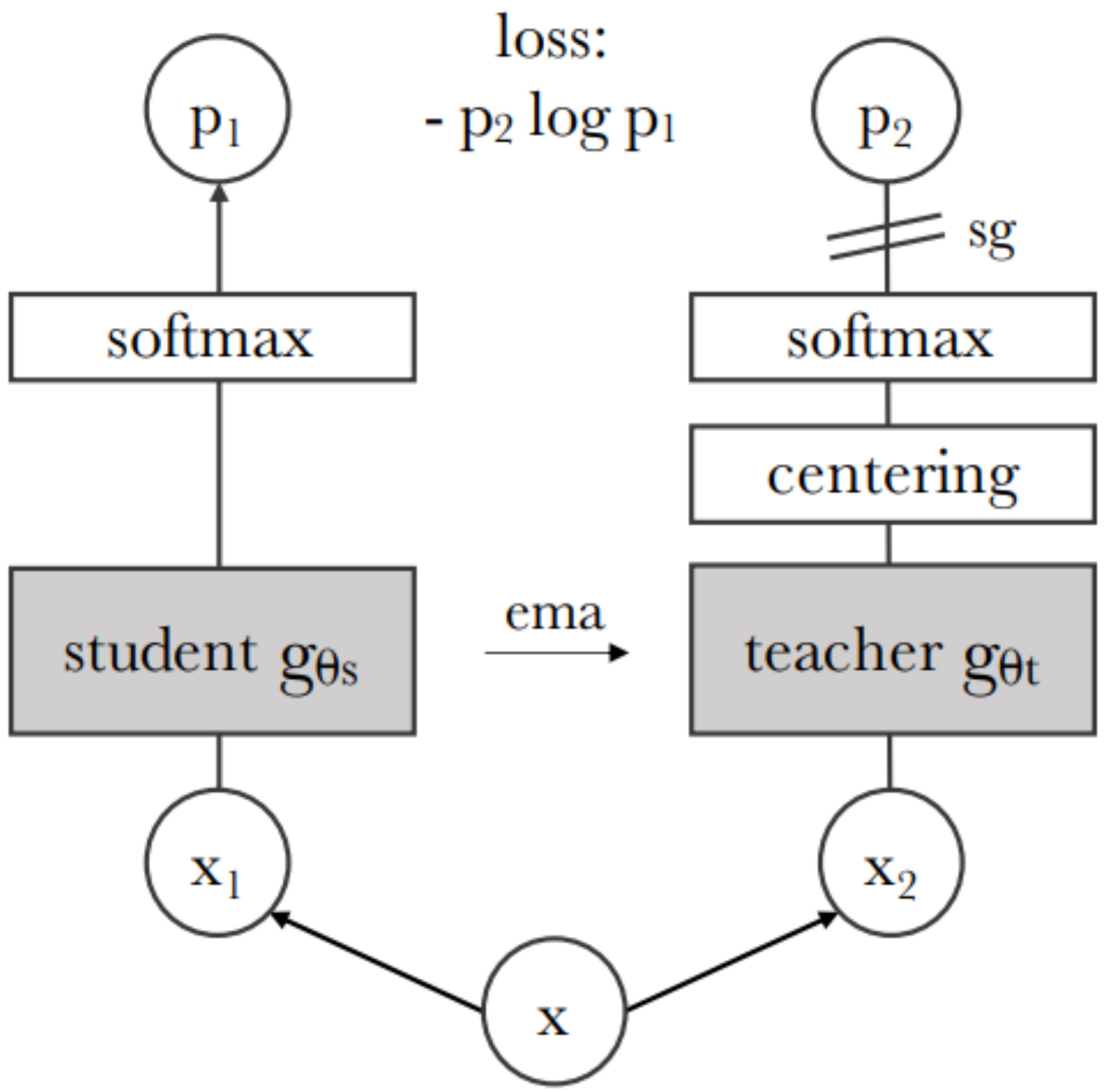
(the normalized temperature-scaled cross-entropy loss)

$$\ell_{i,j} = -\log \frac{\exp(\mathrm{sim}(\boldsymbol{z}_i, \boldsymbol{z}_j)/\tau)}{\sum_{k=1}^{2N} \mathbb{1}_{[k \neq i]} \exp(\mathrm{sim}(\boldsymbol{z}_i, \boldsymbol{z}_k)/\tau)}$$



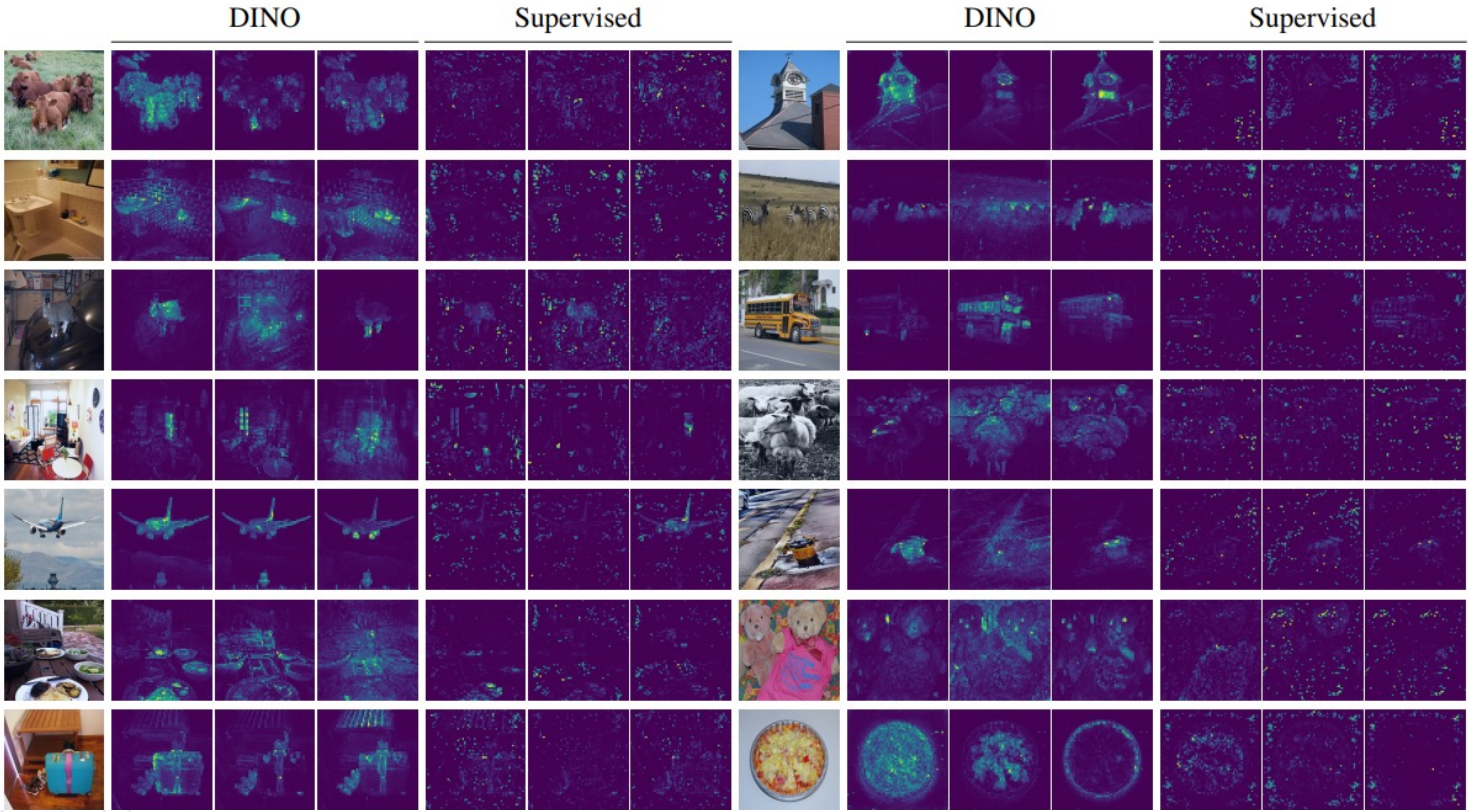
A simple framework for contrastive learning of visual representation, https://arxiv.org/abs/2002.05709

DINO



Emerging Properties in Self-Supervised Vision Transformers

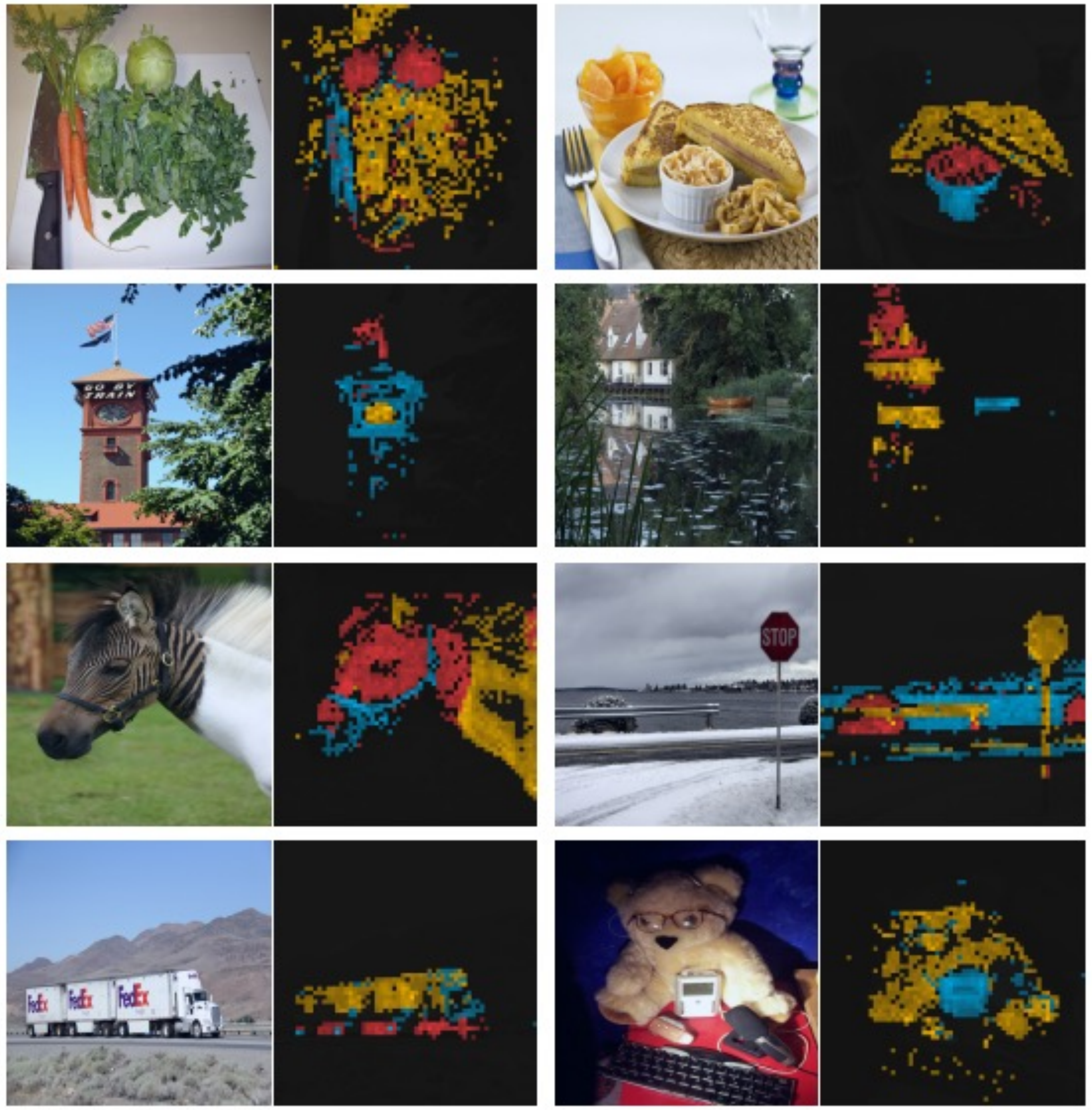
Attention maps visualization

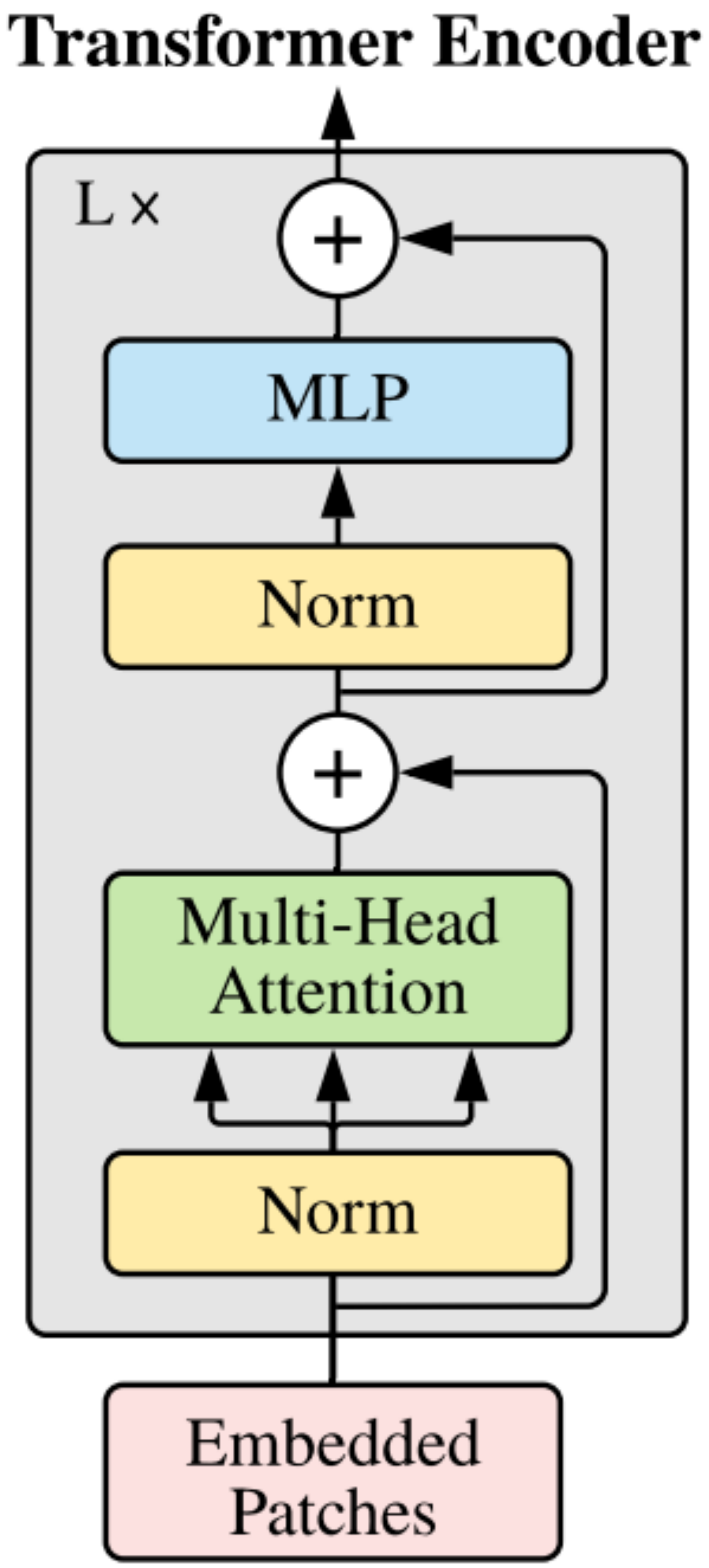


Attention maps visualization

https://ai.meta.com/blog/dino-paws-computer-vision-with-self-supervised-transformers-and-10x-more-efficient-training/

Attention maps from **multiple heads**





Takeaways

1. **No Label Dependency:** Unsupervised learning doesn't rely on labeled data. It aims to learn the underlying structure or distribution of the data without using explicit annotations or feedback.
2. **Data Exploration:** It's excellent for exploratory analysis to uncover hidden patterns, groupings, or relationships in datasets.
3. **Feature Extraction and Representation Learning:** Unsupervised methods, especially deep learning ones like autoencoders, are powerful for representation learning, capturing meaningful and compressed representations of the data
4. **Real-world Relevance:** Many real-world scenarios don't have access to labeled data, making unsupervised methods not just relevant but necessary. For example, understanding customer browsing behavior on a website might involve unsupervised techniques.