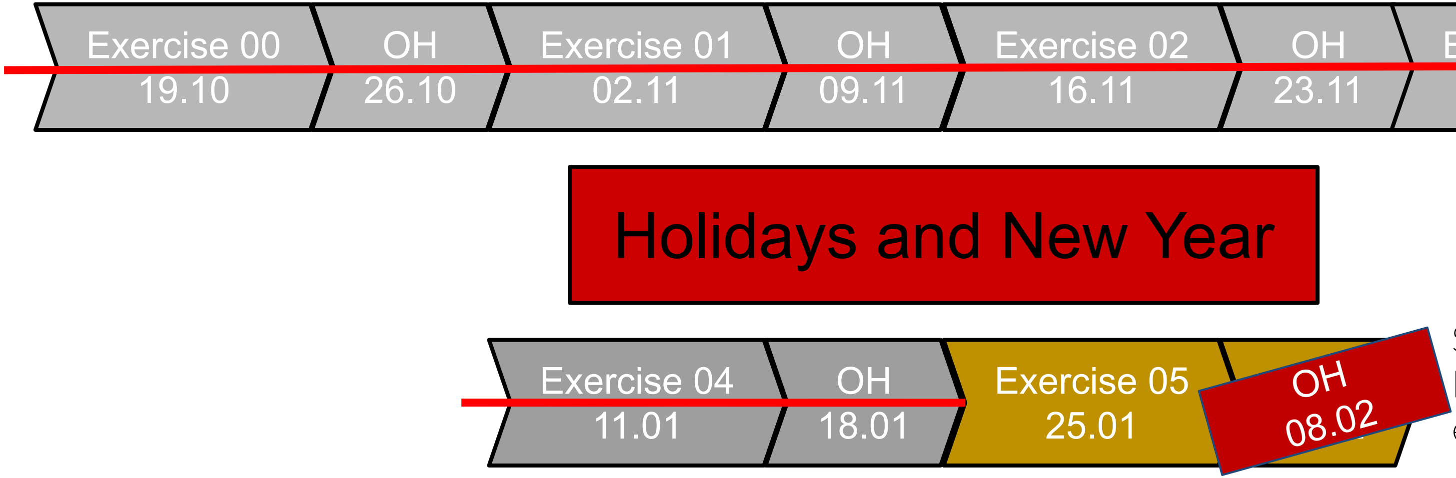
Exercise 5

Exercise 4

- 1 week extension due to submission problems
- please download the exercise folder again if you could not complete the exercise
- please reevaluate the exercise if you have gotten a low score of around 60/100 points

About the Exercise Session

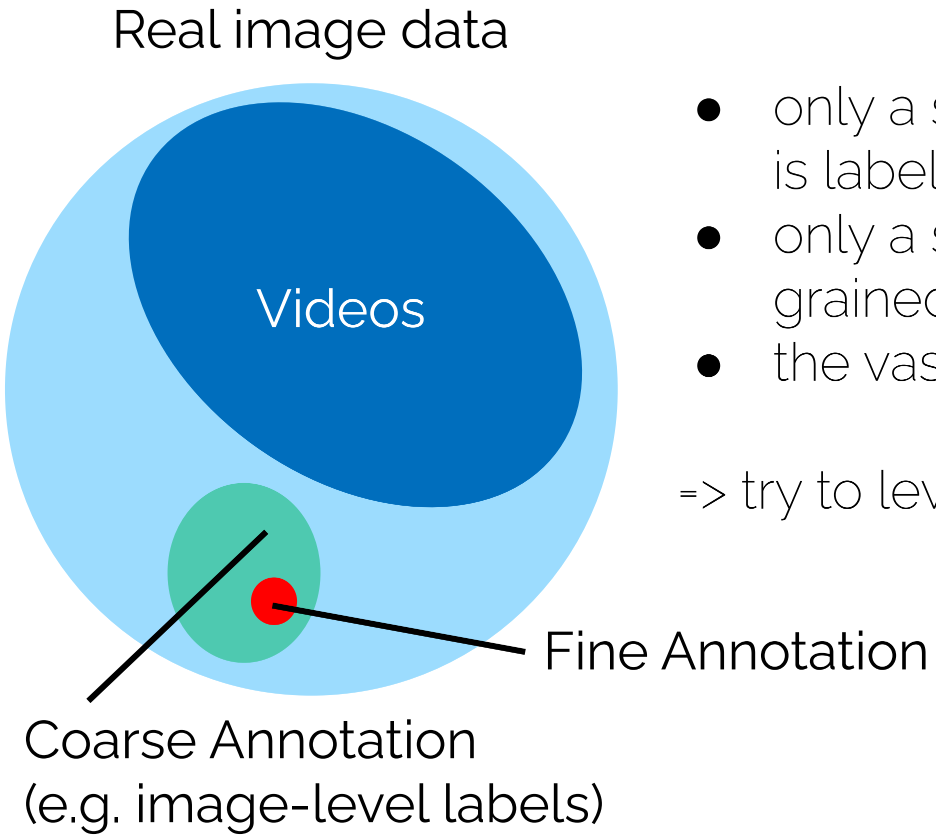
- 2 weeks for each exercise + Office hours (OH) for questions in between



Submission for Exercise 05 also extended by 1 week

Deadline always 23:59 CET on due date

Recap: Unsupervised Segmentation

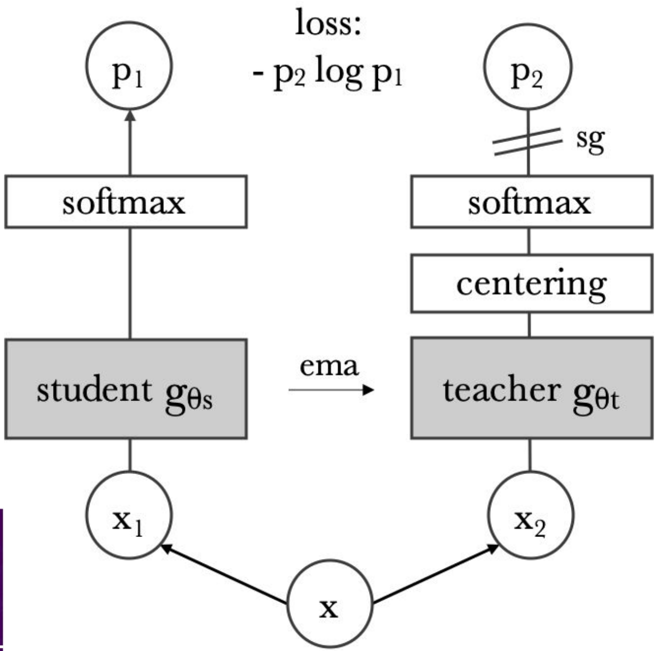


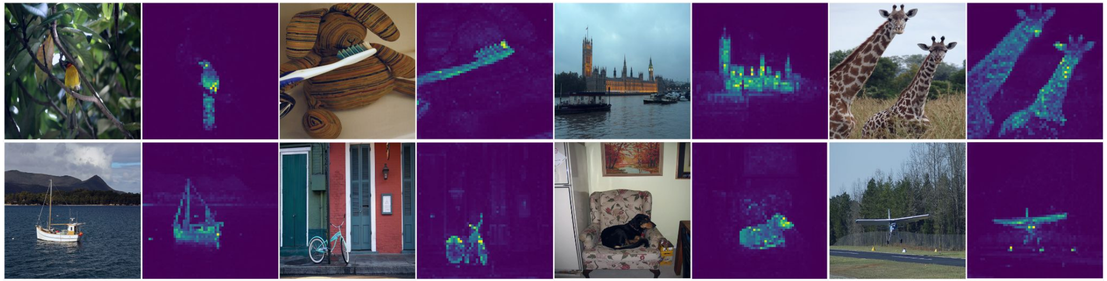
- only a small part of the available image data is labelled
- only a small part of the labelled data has fine grained (pixel-level) annotations
- the vast majority of data is unlabelled

=> try to leverage all available data for your task

Recap: DINO

- You already used pretrained unsupervised (self-supervised) features
- get good results with little training effort on downstream tasks



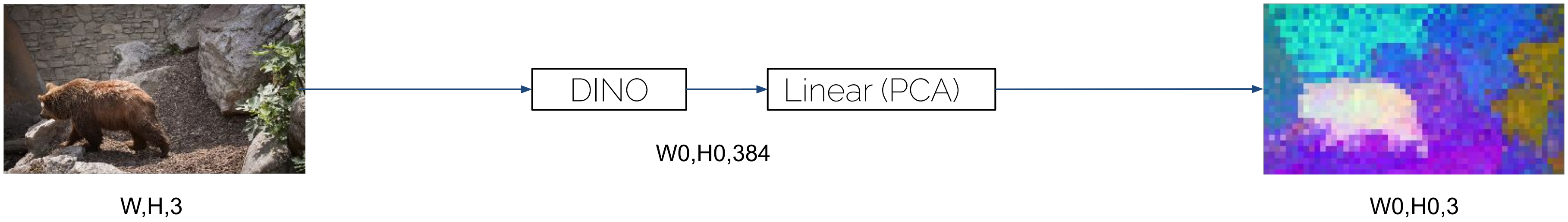


Caron et al. “Emerging Properties in Self-Supervised Vision Transformers”

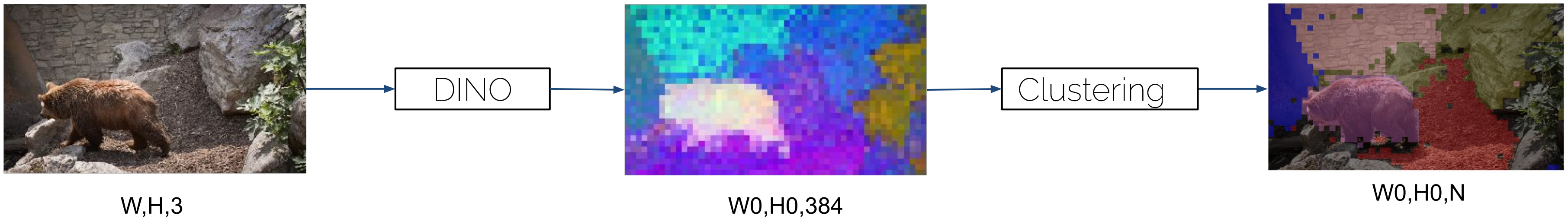
Exercise 5

- **Object Segmentation via Clustering algorithms**
- No supervision available
- Pretrained powerful image embeddings (learned with self-supervision)
- Comparing different methods for unsupervised clustering
  - k-Means
  - Gaussian Mixture Models
  - Spectral Clustering

DINO: Self Supervised Learning



Exercise 5: No Supervision



k-M

$$E(\{\boldsymbol{\mu}_1, \dots \boldsymbol{\mu}_K\}, \boldsymbol{Z}) = \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik} d(\boldsymbol{x}_i, \boldsymbol{\mu}_k)$$

$$\min_{\boldsymbol{Z}} E(\{\boldsymbol{\mu}_1, \dots \boldsymbol{\mu}_K\}, \boldsymbol{Z})$$

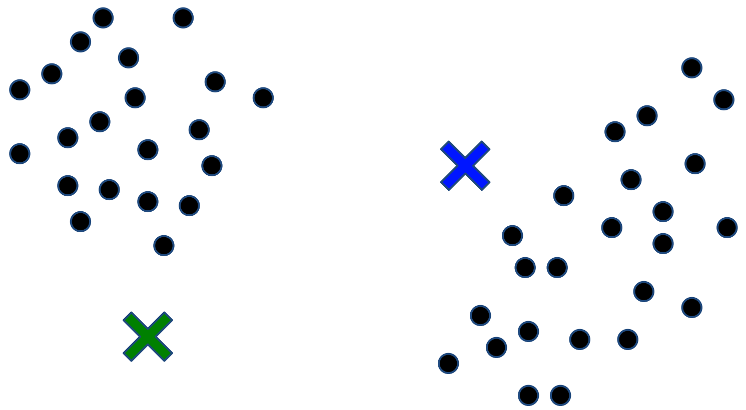
$$z_i k = \begin{cases} 1 & \text{if } k = \arg\min_j \|\boldsymbol{x}_i - \boldsymbol{\mu}_j\|^2 \\ 0 & \text{else.} \end{cases}$$

$$\boldsymbol{\mu}_k = \frac{1}{N_k} \sum_{i=1}^{N} z_{ik} \boldsymbol{x}_i, \quad \text{where } N_k = \sum_{i=1}^{N} z_{ik}$$

$$D_i^2 = \min\{\|\boldsymbol{x}_i - \boldsymbol{\mu}_1\|^2, \dots, \|\boldsymbol{x}_i - \boldsymbol{\mu}_k\|^2\}$$

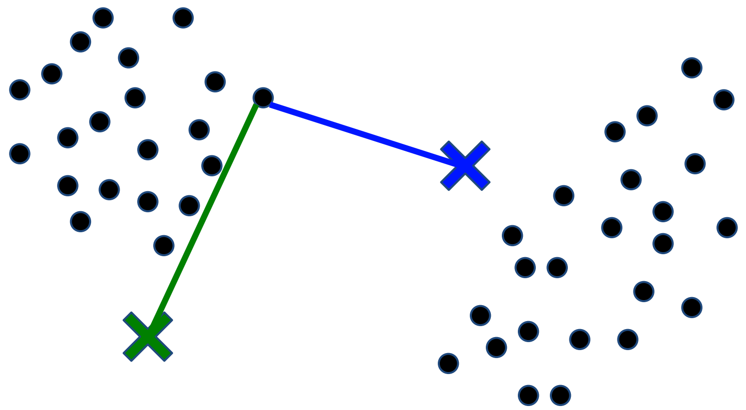
k-Means Clustering

0. initialize the centers



k-Means Clustering

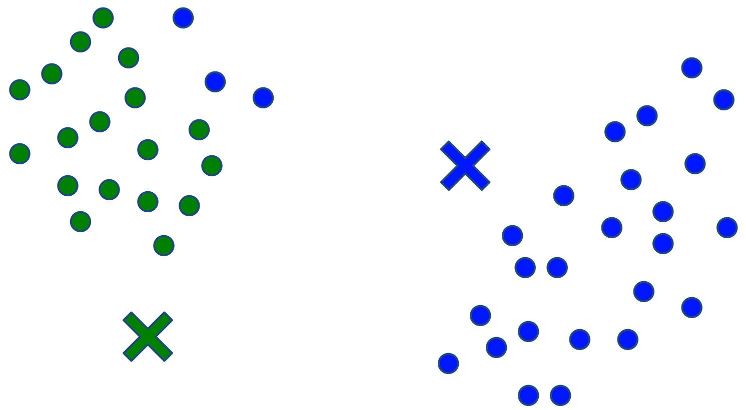
1. calculate the distance to the centers



e.g. $D_i^2 = \min\{\|\boldsymbol{x}_i - \boldsymbol{\mu}_1\|^2, \ldots, \|\boldsymbol{x}_i - \boldsymbol{\mu}_k\|^2\}$

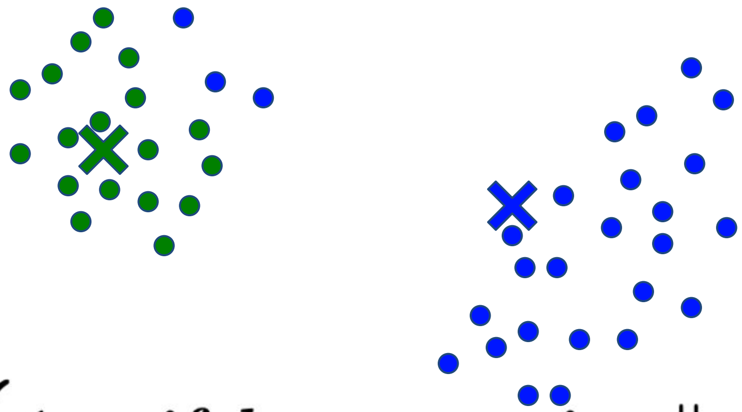
k-Means Clustering

2. assign datapoints to centers based on distance



k-Means Clustering

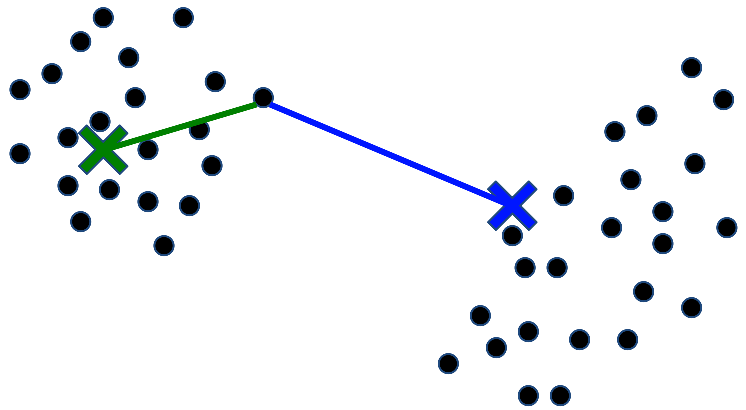
3. calculate new centers



e.g. $$z_{ik} = \begin{cases} 1 & \text{if } k = \arg\min_j \|\boldsymbol{x}_i - \boldsymbol{\mu}_j\|^2 \\ 0 & \text{else.} \end{cases}$$

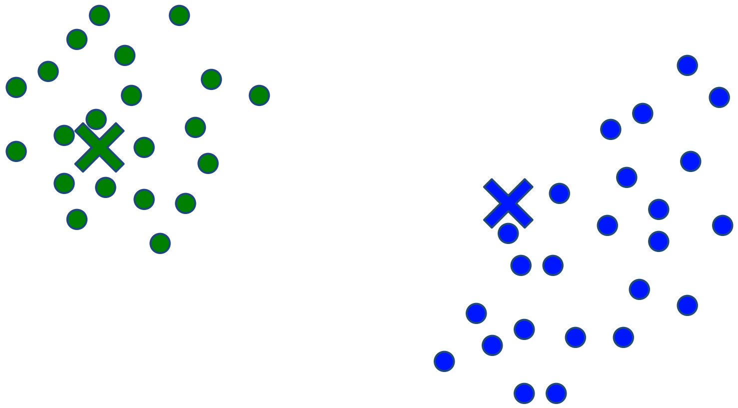
k-Means Clustering

1. calculate the distance to the centers



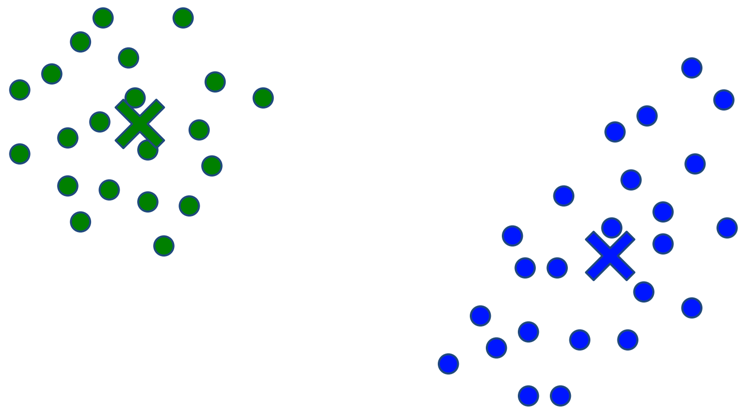
k-Means Clustering

2. assign datapoints to centers based on distance



k-Means Clustering

3. calculate new centers



k-Means in formulas

Cost Function
$$E(\{\boldsymbol{\mu}_1, \ldots \boldsymbol{\mu}_K\}, \boldsymbol{Z}) = \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik} d(\boldsymbol{x}_i, \boldsymbol{\mu}_k)$$

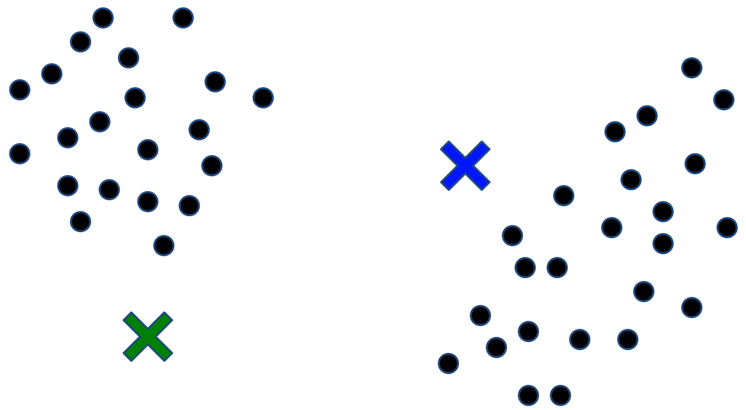
Objective
$$\min_{\boldsymbol{Z}} E(\{\boldsymbol{\mu}_1, \ldots \boldsymbol{\mu}_K\}, \boldsymbol{Z})$$

1. Assign points to clusters
$$z_{ik} = \begin{cases} 1 & \text{if } k = \arg\min_j \|\boldsymbol{x}_i - \boldsymbol{\mu}_j\|^2 \\ 0 & \text{else.} \end{cases}$$

2. Recalculate centroids
$$\boldsymbol{\mu}_k = \frac{1}{N_k} \sum_{i=1}^{N} z_{ik} \boldsymbol{x}_i, \quad \text{where } N_k = \sum_{i=1}^{N} z_{ik}$$

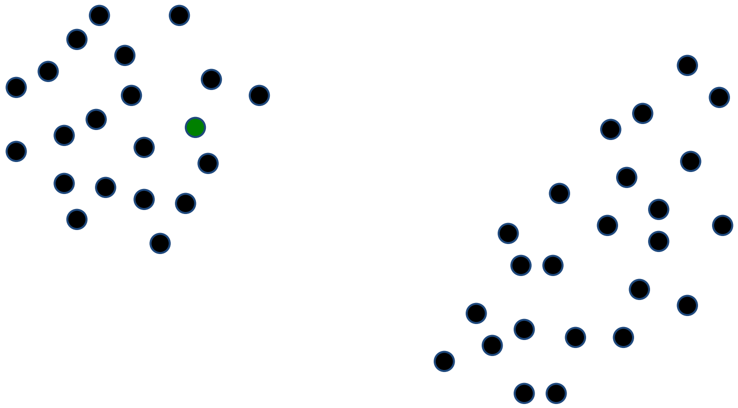
k-Means++

How to initialize the centers?



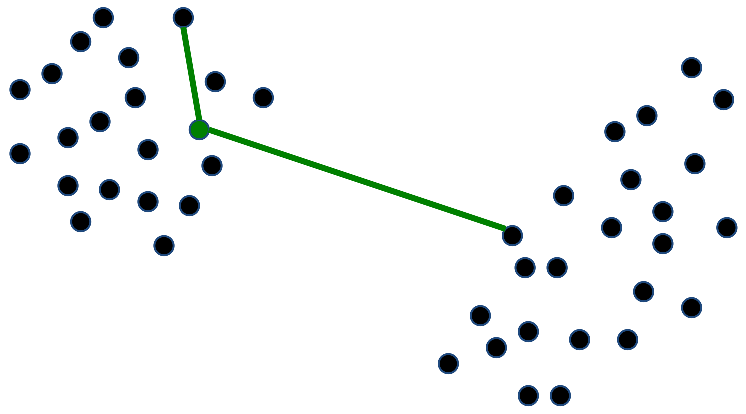
k-Means++

1. choose a random datapoint as the first center



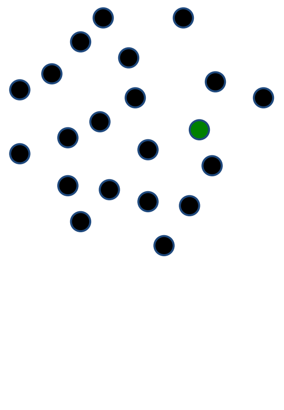
k-Means++

2. compute the minimum distance of all datapoint to all centers



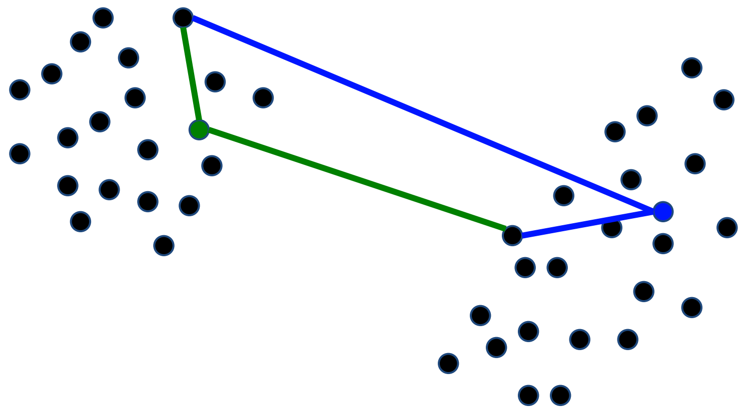
k-Means++

3. choose the next centroid from a categorical sampling



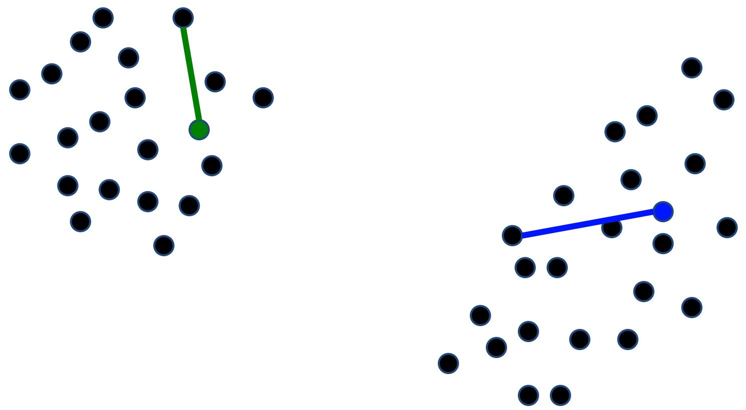
k-Means++

2. compute the minimum distance of all datapoint to all centers



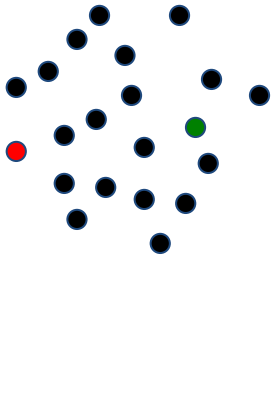
k-Means++

2. compute the minimum distance of all datapoint to all centers



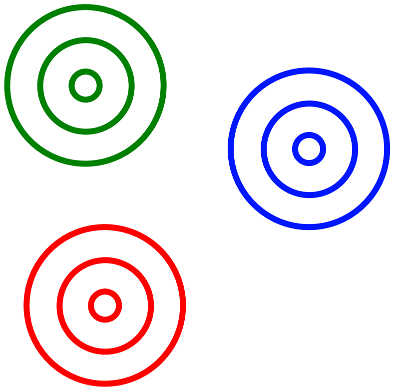
k-Means++

3. choose the next centroid from a categorical sampling

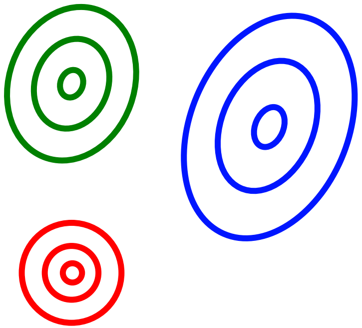


Gaussian Mixture Models

- K-Means is limited in its expressiveness
- Can we add additional degrees of freedom to our model to more accurately represent our data?



k-Means



Gaussian Mixture Model

Difference

- k-Means models the probability of a datapoint belonging to a class only on the centroids of the clusters
- Gaussian Mixture Models model the probability as

$$p(z_{ik} = 1) = \frac{\pi_k \mathcal{N}(\boldsymbol{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\boldsymbol{x}_i | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}$$

- GMMs are equivalent to k-Means if all responsibilities are equal and the covariances are identity matrices

GMMs calculate the parameters

- GMMs are commonly solved by using the expectation-maximization algorithm
- E-step

$$p(z_{ik}=1)=\frac{\pi_k^{(t)}\mathcal{N}(\boldsymbol{x}_i|\boldsymbol{\mu}_k^{(t)},\boldsymbol{\Sigma}_k^{(t)})}{\sum_{j=1}^{K}\pi_j^{(t)}\mathcal{N}(\boldsymbol{x}_i|\boldsymbol{\mu}_j^{(t)},\boldsymbol{\Sigma}_j^{(t)})}$$

Responsibilities

- M-step

$$N_k=\sum_{i=1}^{N}p(z_{ik}=1)$$

$$\boldsymbol{\mu}_k^{(t+1)}=\frac{1}{N_k}\sum_{i=1}^{N}p(z_{ik}=1)\boldsymbol{x}_i$$

Centroids

$$\boldsymbol{\Sigma}_k^{(t+1)}=\frac{1}{N_k}\sum_{i=1}^{N}p(z_{ik}=1)(\boldsymbol{x}_i-\boldsymbol{\mu}_k^{(t+1)})(\boldsymbol{x}_i-\boldsymbol{\mu}_k^{(t+1)})^{\top}$$

Covariances

$$\pi_k^{(t+1)}=\frac{N_k}{N}$$

Mixing Coefficients

Spectral Clustering

- spectral clustering is a unsupervised segmentation methods build upon spectral graph theory
- given a undirected graph

$$G = (V, E)$$

with its adjacency matrix

$$W = w(u, v) : (u, v) \in E$$

we can segment the graph based on the eigenvalues and eigenvectors of the Laplacian

$L = D - W$ where D is the row-wise sum of W

How to cluster?

- foreground-background segmentation

  1. Calculate the eigenvectors of the Laplacian
  2. Select the **second** eigenvector
  3. Threshold the eigenvector

- there are more ways to cluster for
  - object localization
  - multi-class segmentation

  We refer you to the original paper/code for more details

But we are working with images?

- Construct an adjacency matrix from the image
- We follow [Deep Spectral Methods]

$$W_{\text{knn}}(u,v) = \begin{cases} 1 - \|\psi(u) - \psi(v)\|, & u \in \text{KNN}_{\psi}(v), \\ 0, & \text{otherwise}, \end{cases}$$

$$W_{\text{feat}} = \boldsymbol{f}\boldsymbol{f}^{\top} \odot (\boldsymbol{f}\boldsymbol{f}^{\top} > 0)$$

$$W = W_{\text{feat}} + \lambda W_{\text{knn}}$$

and have a image- and learning-based part

Melas-Kyriazi et al., Deep Spectral Methods: A Surprisingly Strong Baseline for Unsupervised Semantic Segmentation and Localization, CVPR 2022

Image based adjacency

$$W_{\mathrm{knn}}(u,v) = \begin{cases} 1 - \|\psi(u) - \psi(v)\|, & u \in \mathrm{KNN}_{\psi}(v), \\ 0, & \text{otherwise,} \end{cases}$$

$$\psi(u) = (\underbrace{\cos(c_H(u)), \sin(c_H(u)), c_S(u), c_V(u)}_{\text{color values from the HSV color space}}, \underbrace{p_x(u), p_y(u)}_{\text{pixel position } (p_x, p_y)})$$

Feature based adjacency

$$W_{\mathrm{feat}} = \boldsymbol{f}\boldsymbol{f}^\top \odot (\boldsymbol{f}\boldsymbol{f}^\top > 0)$$

$\boldsymbol{f}$ deep image feature e.g. DINO

- as the deep image features restrict the resolution, we have to adapt the resolution of the image-based adjacency
- this also reduces the size of $\boldsymbol{W}$ significantly
- additional post processing steps (e.g. CRFs) increases resolution and can improve results even further

Downsides

- low resolution of the segmentation without post-processing
- k-Means and GMMs are initialization dependent
  => k-means++ does not solve it completely
- number of clusters is a hyperparameter to tune
  => spectral clustering is able to adapt on the fly

Links

- Test server: https://cv3dst.cvai.cit.tum.de/login
- If you have trouble registering https://forms.gle/yZkZiDiyHxWuNqQG7
- Data for Exercise 04: https://vision.in.tum.de/webshare/g/cv3dst/exercise_05.zip

Links for the individual datasets

- MOT
  https://vision.in.tum.de/webshare/g/cv3dst/datasets/MOT16.zip
- market
  https://vision.in.tum.de/webshare/g/cv3dst/datasets/market.zip
- obj_seg
  https://vision.in.tum.de/webshare/g/cv3dst/datasets/obj_seg.zip
- reid_gnn
  https://vision.in.tum.de/webshare/g/cv3dst/datasets/reid_gnn.zip