Photogrammetry & Robotics Lab

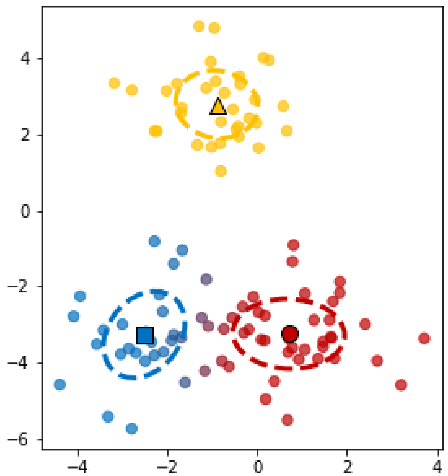
Machine Learning for Robotics and Computer Vision Tutorial

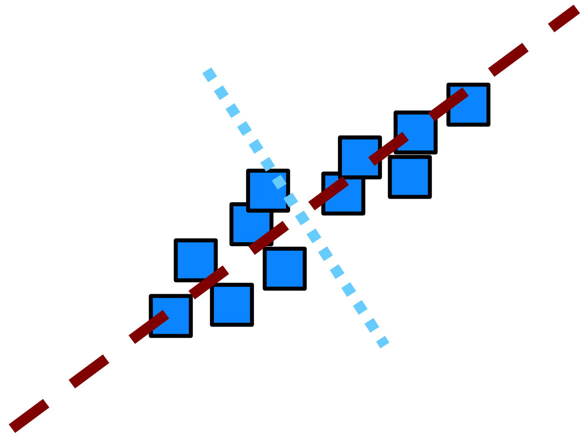
Recap Unsupervised Learning & ML for CV Tasks

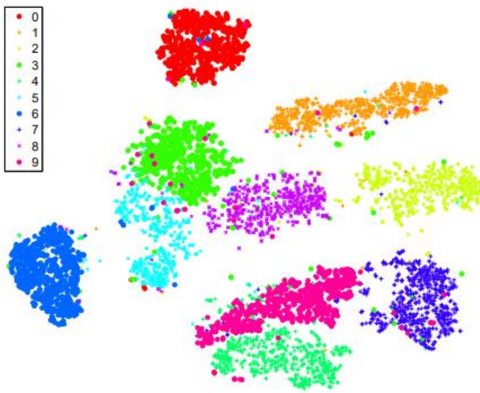
Jens Behley

Unsupervised Learning

Recap: Unsupervised Learning

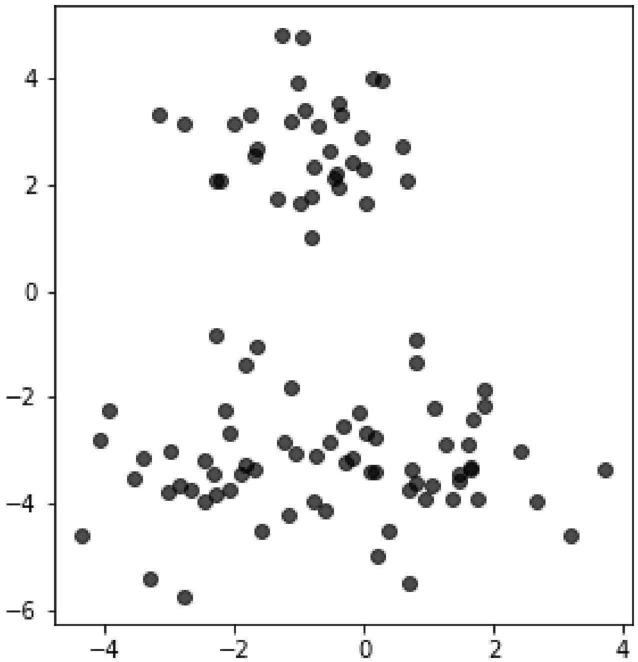






- Discussed several unsupervised learning approaches solving different tasks:
  - Density Estimation (Gaussian Mixture Models)
  - Dimensionality Reduction (PCA)
  - Visualization (t-SNE)

Recap: Unsupervised Learning



- Unsupervised setting, we have data points without specific labels
- **Goal:** Learn a representation of the data that can be used to extract information or gain insights

Recap: Latent variables

- Assume that data can be decomposed in multiple parts by a **latent variable** h:

$$P(\mathbf{x}) = \sum_{i=0}^{K} P(h = i, \mathbf{x})$$

$$= \sum_{k=1}^{K} P(h = k) P(\mathbf{x} | h = k)$$

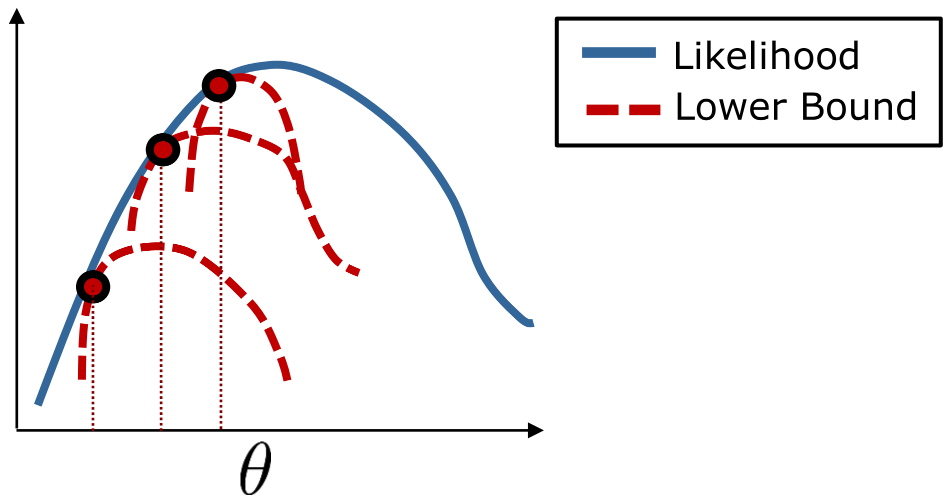
- For the **Gaussian Mixture Model (GMM)**:

$$P(h = k) = \lambda_k \qquad \text{with} \qquad \sum_k \lambda_k = 1$$

$$P(\mathbf{x} | h = k) = \mathcal{N}(\mathbf{x} | \mu_k, \Sigma_k)$$

[CVMLI Chap. 7] [MML Chap. 11]

Recap:Expectation Maximization



- Instead of maximizing directly the log-likelihood, we maximize a lower bound with the same parameters
- Alternate between getting a new lower bound (E-Step) and maximizing the lower bound (M-Step)

[CVMLI Chap. 7] [MML Chap. 11]

EM for Gaussian Mixture Model

- Algorithm for learning a GMM:

1. Initialize $\mu_k^{[0]}$ by selecting K random examples $\mathbf{x} \in \mathcal{X}$, $\Sigma_k^{[0]} = \mathbf{I}$, and $\lambda_k^{[0]} = K^{-1}$
2. E-Step: Determine $r_{ik}^{[t-1]}$ by computing:

$$r_{ki}^{[t-1]} = P(h = k|\mathbf{x}_i) = \frac{\lambda_k^{[t-1]} \mathcal{N}\left(\mathbf{x}_i | \mu_k^{[t-1]}, \Sigma_k^{[t-1]}\right)}{\sum_k \lambda_k^{[t-1]} \mathcal{N}\left(\mathbf{x}_i | \mu_k^{[t-1]}, \Sigma_k^{[t-1]}\right)}$$

3. M-Step: Update parameters:

$$\lambda_k^{[t]} = \frac{1}{N} \sum_{i=1}^{N} r_{ik}^{[t-1]} \qquad \mu_k^{[t]} = \frac{\sum_{i=1}^{N} r_{ik}^{[t-1]} \mathbf{x}_i}{\sum_{i=1}^{N} r_{ik}^{[t-1]}}$$

$$\Sigma_k^{[t]} = \frac{\sum_{i=1}^{N} r_{ik}^{[t-1]} \left(\mathbf{x}_i - \mu_k^{[t]}\right)\left(\mathbf{x}_i - \mu_k^{[t]}\right)^T}{\sum_{i=1}^{N} r_{ik}^{[t-1]}}$$

4. Repeat E-Step & M-Step until convergence.

Relation to k-Means Clustering

- K-Means is special case with hard responsibilities, i.e., 0 or 1, and fixed Σ

- K-Means algorithm:
  1. Initialize means/cluster center randomly.
  2. Assign each point to nearest cluster center.
  3. Update cluster center by mean of assigned points

- Choice of initial cluster centers (more) important.

Recap: PCA

- Algorithm for PCA
  1. **Standardize** the data, i.e., subtract mean and divide each dimension by it's variance
  2. Determine Eigenvectors and Eigenvalues to get basis $\mathbf{B} \in \mathbb{R}^{D \times M}$

- Projected points are then given by

$$\tilde{\mathbf{x}}_* = \mathbf{B}\mathbf{B}^T \mathbf{x}_*$$

  where

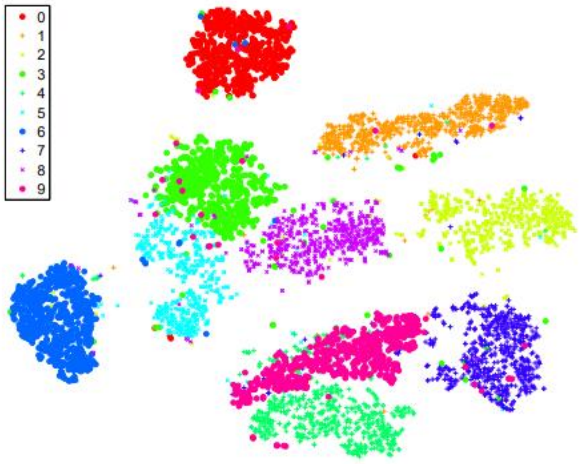
$$x_*^{(d)} = \frac{x^{(d)} - \mu^{(d)}}{\sigma_d}$$

- Reconstruction in original space by de-standardization of $\tilde{\mathbf{x}}_*$

[MML Chap. 10]

Recap: Visualizing high-dimensional data

MNIST dataset
28x28 Images



Low-dimensional Embedding via t-SNE

- Here, want to find low-dimensional embedding that retains distances of high-dimensional data

[van der Maaten, 2008]

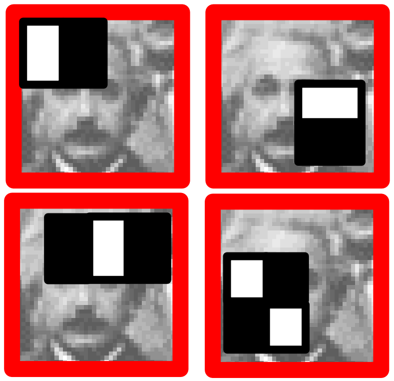
T-SNE Implementation

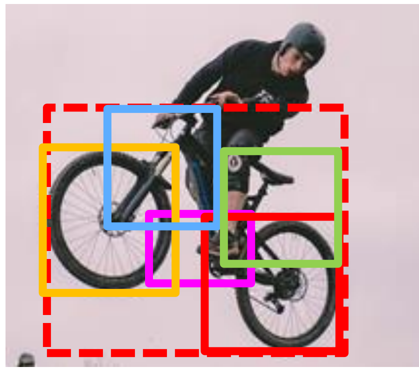
- Implementation: `sklearn.manifold.TSNE`
- Nice visualization also of the learning process: https://distill.pub/2016/misread-tsne/

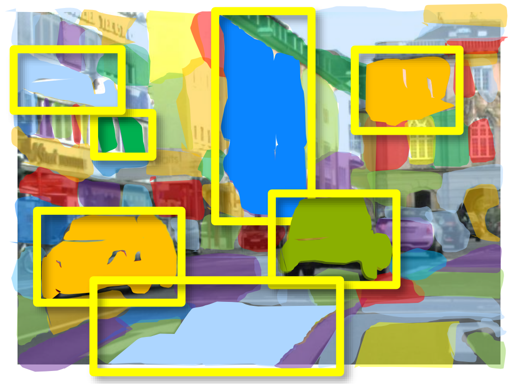
Questions?

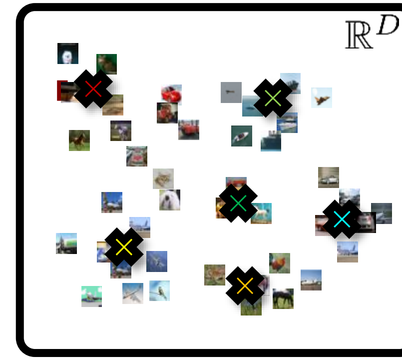
ML for CV Tasks

Historical Overview of Methods



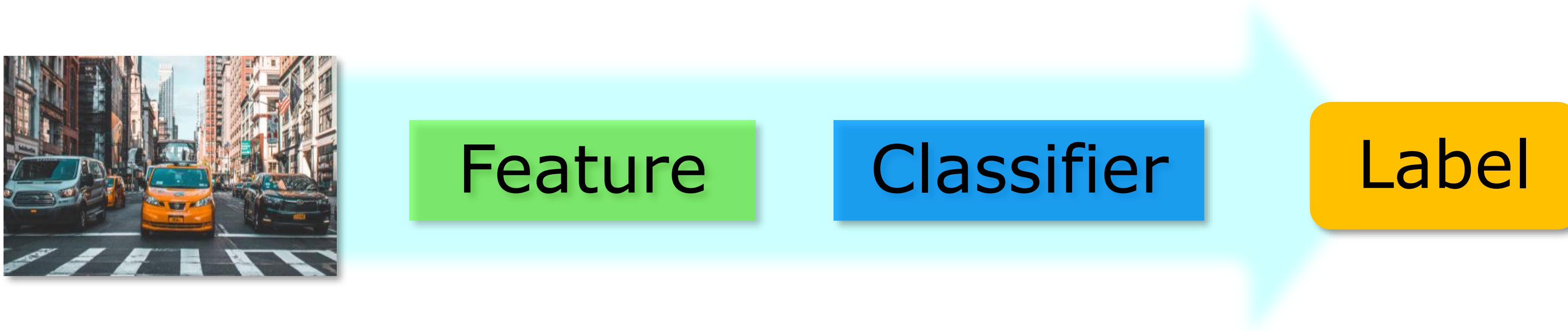






- Even though most methods are replaced by better variants, having an understanding what happened before important
- Research is kind of incremental: “Standing on the shoulders of giants”

Recap: Feature Engineering



- Applications to Computer Vision tasks: Extract features and apply supervised learning methods
- Most of the time: designing task-specific features → **feature engineering**

Covered Methods (Timeline)

- Viola Jones Detector (2001)
- Bag-of-Words (2004)
- HoG descriptor (2005)
- Spatial Pyramids (2006)
- Deformable Part Models (2009)
- Selective Search (2011)

- Datasets:
  - Pascal VOC (2007-2012)
  - ImageNet (2009/2010)

Main ML Conferences

- Conference on Neural Information Processing Systems (NeurIPS)
- International Conference on Machine Learning (ICML)
- International Conference on Learning Representations (ICLR)

Main Vision Conferences

- IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)
- IEEE/CVF International Conference on Computer Vision Conference (ICCV)
- European Conference on Computer Vision (ECCV)

Main Robotics Conferences

- IEEE International Conference on Robotics and Autonomation (ICRA)
- IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
- Robotics: Science & Systems (RSS)

Questions?

See you next week!