CS 446/ECE 449: Machine Learning

A. G. Schwing

University of Illinois at Urbana-Champaign, 2020

L21: k-Means Clustering

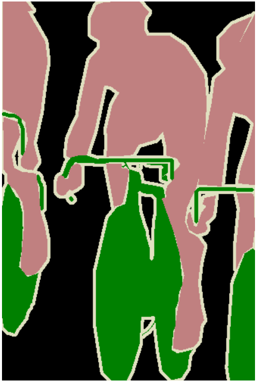
Goals of this lecture

- Understanding clustering concept
- Getting to know k-Means

Reading material:

- K. Murphy; Machine Learning: A Probabilistic Perspective; Chapter 11

Recap: Semantic Image Segmentation

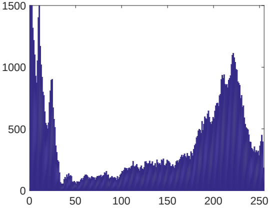


How did we do it?

What if we don't have labels?

Image

Intensities



Clustering

Group perceptually or according to another metric similar regions

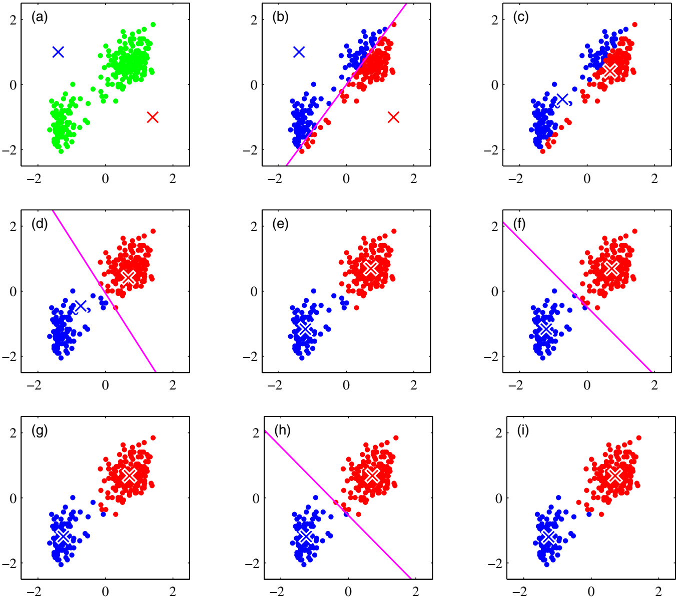
How can we do it?

Video

kMeans/Lloyd's Algorithm (informal):

- Initialize: pick K random points as cluster centers μ_k
- Iterate:
  - Assign data points $x^{(i)}$ to closest cluster center according to some metric
  - Update the cluster center to be the average of its assigned points
  - Stopping criterion: when no points' assignments change

2D Example:



Formal description:

What cost function does kMeans optimize? (distortion measure)

$$\min_{\mu} \min_{r} \sum_{i\in\mathcal{D}} \sum_{k=1}^{K} \frac{1}{2} r_{ik} \|x^{(i)} - \mu_k\|_2^2 \quad \text{s.t.} \quad \left\{ \begin{array}{ll} r_{ik} \in \{0,1\} & \forall i,k \\ \sum_{k=1}^{K} r_{ik} = 1 & \forall i \end{array} \right.$$

What does the constraint remind us of?

How to optimize this cost function?

Cost function:

$$\min_{\mu} \min_{r} \sum_{i \in \mathcal{D}} \sum_{k=1}^{K} \frac{1}{2} r_{ik} \|x^{(i)} - \mu_k\|_2^2 \quad \text{s.t.} \quad \left\{ \begin{array}{ll} r_{ik} \in \{0, 1\} & \forall i, k \\ \sum_{k=1}^{K} r_{ik} = 1 & \forall i \end{array} \right.$$

Alternate optimization:

- Optimize for r given μ

$$r_{ik} = \left\{ \begin{array}{ll} 1 & \text{if } k = \arg\min_{k \in \{1,\ldots,K\}} \|x^{(i)} - \mu_k\|_2^2 \\ 0 & \text{otherwise} \end{array} \right.$$

- Optimize for μ given r

$$\nabla_{\mu_k}: \qquad \sum_{i \in \mathcal{D}} r_{ik}(x^{(i)} - \mu_k) = 0$$

$$\mu_k = \frac{\sum_{i \in \mathcal{D}} r_{ik} x^{(i)}}{\sum_{i \in \mathcal{D}} r_{ik}}$$

Properties of Algorithm:

- Local optimum is found
- Guaranteed to converge in a finite number of iterations
- Running time per iteration:
  - Assign data points to closest cluster center

$$O(KNd)$$

  - Change the cluster center to the average of its assigned points

$$\cancel{O(N)} \quad Nd$$

Extensions:

- kMeans is very sensitive to initialization: kMeans++
- kMeans depends on the feature space: Kernels
- Evaluation

kMeans++: How to initialize kMeans

- Randomly choose first center
- Pick new center with probability proportional to $\|x^{(i)} - \mu_k\|_2^2$ (contribution of $x^{(i)}$ to total error)
- Repeat until K centers are chosen

Try multiple initializations and choose the best

Distance measure:

- Euclidean (most commonly used)
- Cosine
- Non-linear (via Kernels): $\phi(x^{(i)})$

How to evaluate clusters?

- Generative: How well are points reconstructed from the clusters

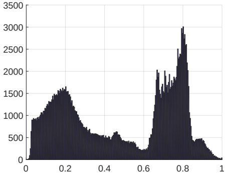
  Distortion

- Discriminative: How well do the clusters correspond to labels

  Purity

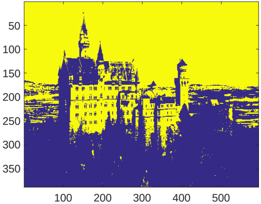
Applications

- Clustering
- Super-pixel segmentation

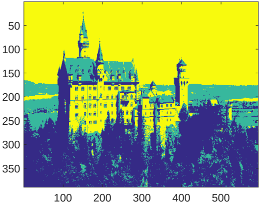


- Grayscale $x^{(i)} \in \mathbb{R}$

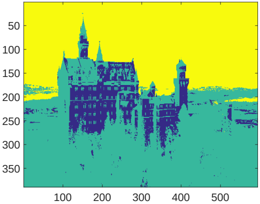
2 clusters



3 clusters



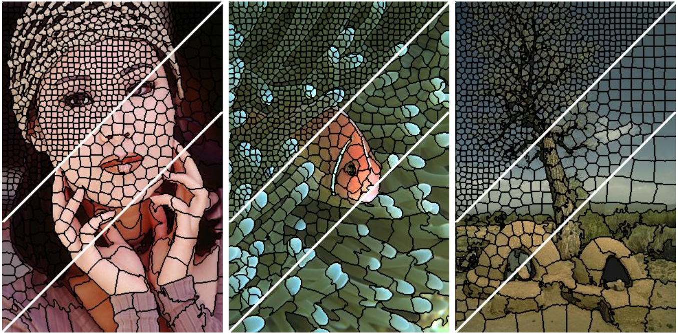
- Color space $x^{(i)} \in \mathbb{R}^3$



How to obtain spatial smoothness?

Augment the feature space $\phi(x^{(i)})$ to contain spatial coordinates in addition to intensities

Superpixel segmentation



Superpixel segmentation

- 5D feature space (lab color space, 2d spatial coordinates)
- Distance metric treats color dimensions and spatial dimensions differently
- Initial cluster centers are spaced regularly on the image and slightly perturbed to avoid edges

Summary:

Pros:

- Simple
- Easy to implement

Cons:

- Need to choose K
- Sensitive to outliers
- Can get stuck in local minima
- All cluster centers have same parameters (non-adaptive)
- Can be slow $O(KNd)$

Quiz:

- What is the cost function for kMeans?
- What are the steps of the kMeans algorithm?
- What are the guarantees of the kMeans algorithm?

Important topics of this lecture

- Understanding kMeans
- Getting to know different mechanisms to adjust kMeans