Introduction to intelligent systems

Image processing

Mikkel N. Schmidt

Technical University of Denmark,
DTU Compute, Department of Applied Mathematics and Computer Science.

Overview

1. K-means clustering
2. Data geometry: Image data
3. Technical writing
4. Tasks

Feedback group

- Carl Borg
- Magnus Nordtorp Mabeck
- Aleks Laith Gryn
- Christine Amalie Meinert Cardel

Learning objectives

II Feature normalization and standardization.

II K-means clustering. Model, cost function, parameters, and algorithm.

I Understand the concepts and definitions, and know their application. Reason about the concepts in the context of an example. Use correct technical terminology.

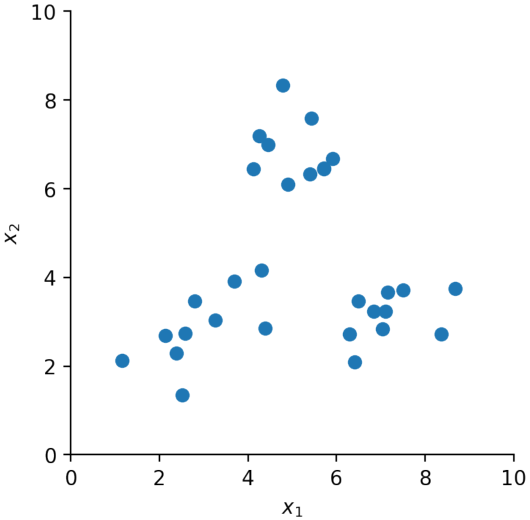
II As above plus: Read, manipulate, and work with technical definitions and expressions (mathematical and Python code). Carry out practical computations. Interpret and evaluate results.

K-means clustering

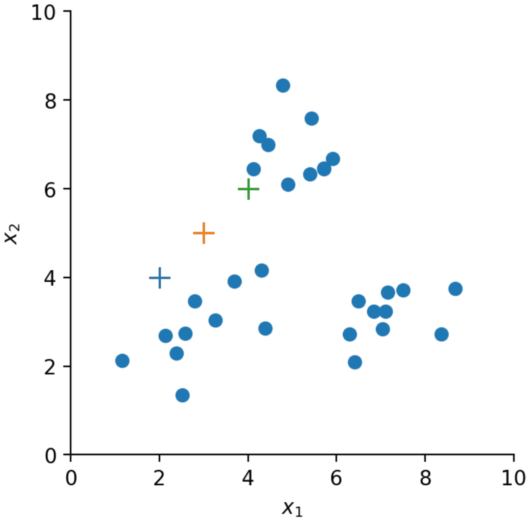
K-means clustering

- Basic idea: Group n observations into K clusters
- Observations belong to the cluster with the closest mean
- The cluster mean serves as a prototype of the cluster

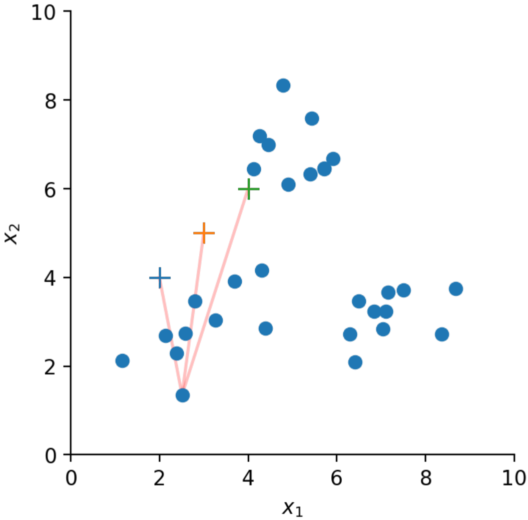
K-means example



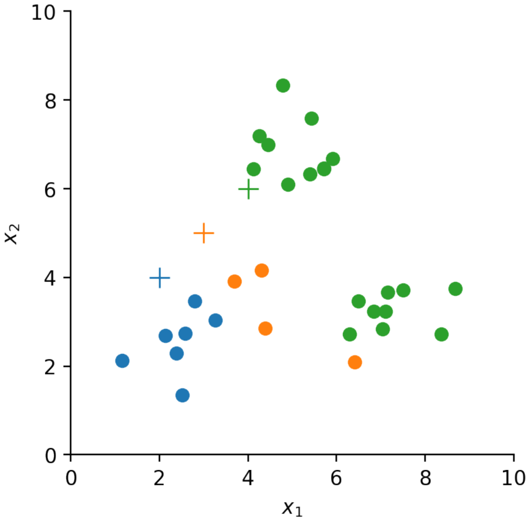
K-means example



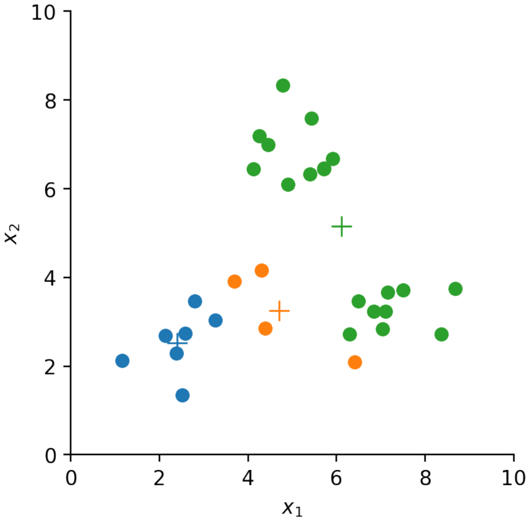
K-means example



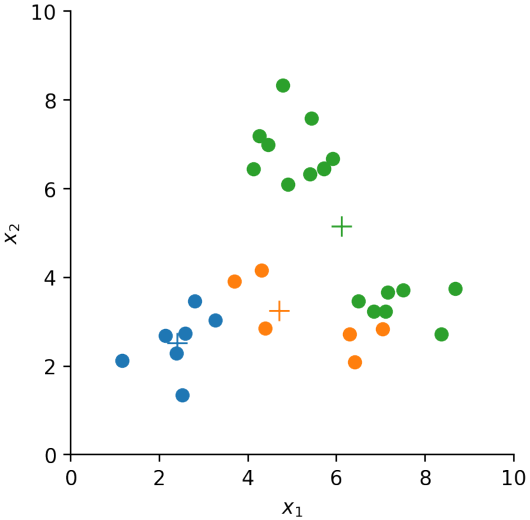
K-means example



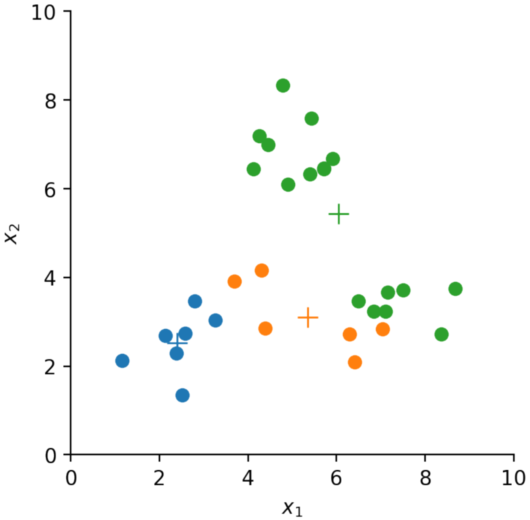
K-means example



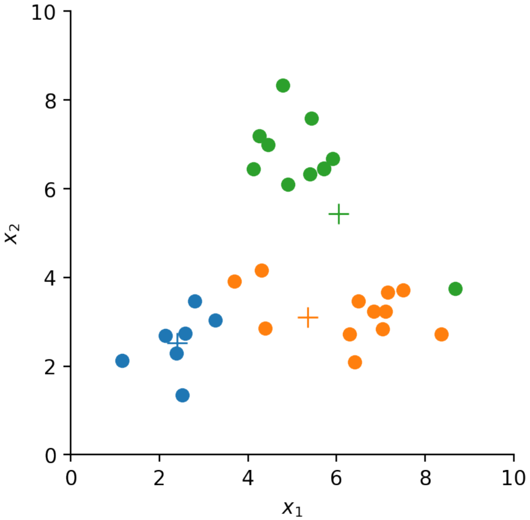
K-means example



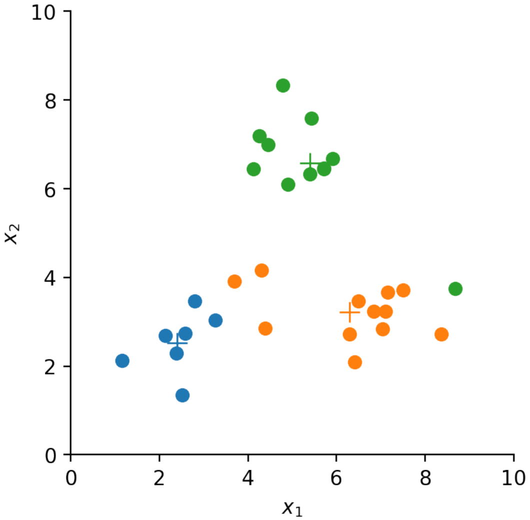
K-means example



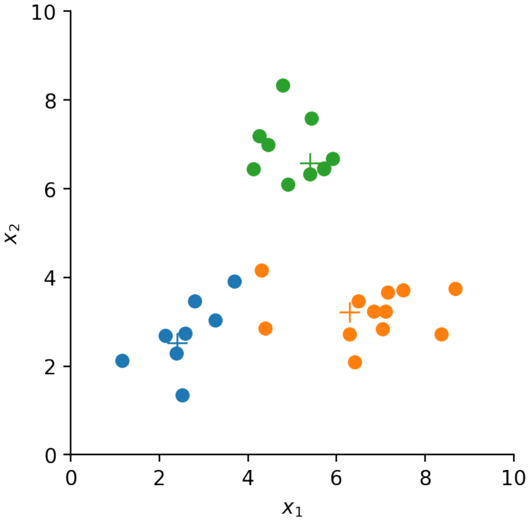
K-means example



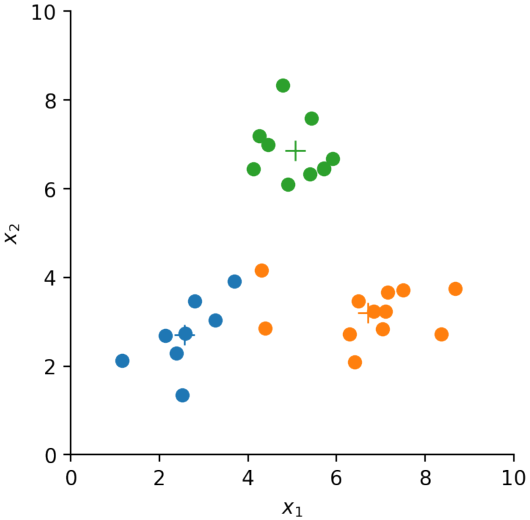
K-means example



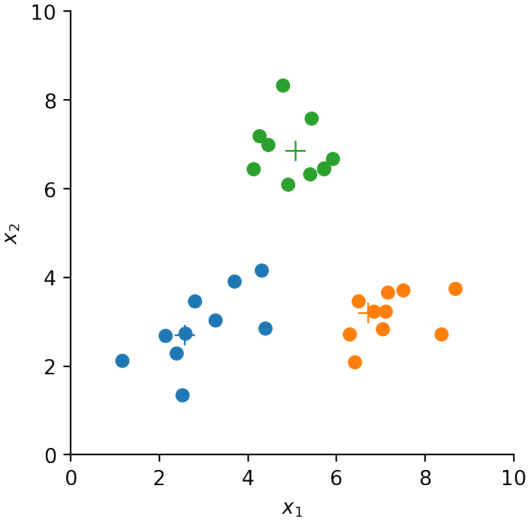
K-means example



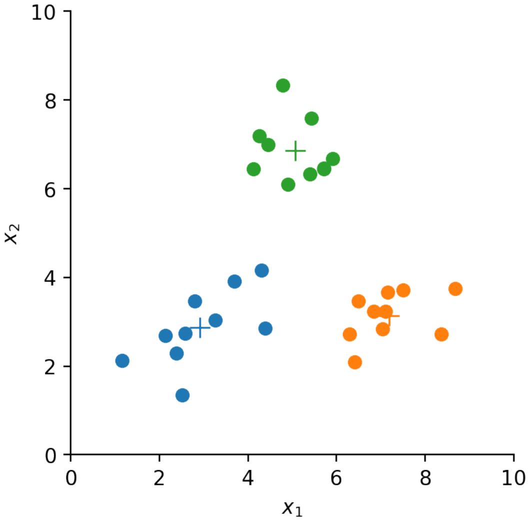
K-means example



K-means example



K-means example



K-means algorithm

Minimization of the objective

$$\underbrace{\min_{\{\boldsymbol{z}_1,\ldots,\boldsymbol{z}_K\}}}_{\substack{\text{Cluster}\\ \text{means}}} \underbrace{\min_{\{\boldsymbol{c}_1,\ldots,\boldsymbol{c}_N\}}}_{\substack{\text{Cluster}\\ \text{assignments}}} \underbrace{\sum_{n=1}^{N} \|\boldsymbol{x}_n - \boldsymbol{z}_{c_n}\|^2}_{\substack{\text{Squared distance}\\ \text{to cluster mean}}}$$

K-means algorithm

Minimization of the objective

$$\underbrace{\min_{\{\boldsymbol{z}_1,\ldots,\boldsymbol{z}_K\}}}_{\substack{\text{Cluster}\\\text{means}}} \underbrace{\min_{\{\boldsymbol{c}_1,\ldots,\boldsymbol{c}_N\}}}_{\substack{\text{Cluster}\\\text{assignments}}} \underbrace{\sum_{n=1}^{N} \|\boldsymbol{x}_n - \boldsymbol{z}_{c_n}\|^2}_{\substack{\text{Squared distance}\\\text{to cluster mean}}}$$

Algorithm

1. Fix cluster means, optimize cluster assignment

$$\min_{\{c_1,\ldots,c_N\}} \sum_{n=1}^{N} \|\boldsymbol{x}_n - \boldsymbol{z}_{c_n}\|^2$$

K-means algorithm

Minimization of the objective

$$\underbrace{\min_{\{\boldsymbol{z}_1,\ldots,\boldsymbol{z}_K\}}}_{\substack{\text{Cluster}\\ \text{means}}} \underbrace{\min_{\{\boldsymbol{c}_1,\ldots,\boldsymbol{c}_N\}}}_{\substack{\text{Cluster}\\ \text{assignments}}} \underbrace{\sum_{n=1}^{N} \|\boldsymbol{x}_n - \boldsymbol{z}_{c_n}\|^2}_{\substack{\text{Squared distance}\\ \text{to cluster mean}}}$$

Algorithm

1. Fix cluster means, optimize cluster assignment

$$\min_{\{c_1,\ldots,c_N\}} \sum_{n=1}^{N} \|\boldsymbol{x}_n - \boldsymbol{z}_{c_n}\|^2$$

2. Fix cluster assignments, optimize cluster means

$$\min_{\{\boldsymbol{z}_1,\ldots,\boldsymbol{z}_K\}} \sum_{n=1}^{N} \|\boldsymbol{x}_n - \boldsymbol{z}_{c_n}\|^2$$

Notation for the objective

The objective can be written as a sum over all data points

$$L = \sum_{n=1}^{N} \left\| \boldsymbol{x}_n - \boldsymbol{z}_{c_n} \right\|^2$$

Notation for the objective

The objective can be written as a sum over all data points

$$L = \sum_{n=1}^{N} \|\boldsymbol{x}_n - \boldsymbol{z}_{c_n}\|^2$$

or as a sum over all data points in each cluster inside a sum over all clusters

$$L = \underbrace{\sum_{k=1}^{K}}_{\text{Clusters}} \quad \underbrace{\sum_{n:c_n=k}}_{\substack{\text{Observations}\\ \text{in cluster } k}} \quad \|\boldsymbol{x}_n - \boldsymbol{z}_k\|^2$$

Notation for the distances

- The squared distance from a data point to a cluster is the squared norm of the difference

$$\left\| \boldsymbol{x}_n - \boldsymbol{z}_k \right\|^2$$

where $\boldsymbol{x}_n$ and $\boldsymbol{z}_k$ are vectors.

Notation for the distances

- The squared distance from a data point to a cluster is the squared norm of the difference

$$\|\boldsymbol{x}_n - \boldsymbol{z}_k\|^2$$

where $\boldsymbol{x}_n$ and $\boldsymbol{z}_k$ are vectors.

- In two dimension, for example, we have

$$\boldsymbol{x}_n = \begin{bmatrix} x_n^{(1)} \\ x_n^{(2)} \end{bmatrix}, \quad \boldsymbol{z}_k = \begin{bmatrix} z_k^{(1)} \\ z_k^{(2)} \end{bmatrix}$$

so the squared distance is given as

$$\|\boldsymbol{x}_n - \boldsymbol{z}_k\|^2 = (x_n^{(1)} - z_k^{(1)})^2 + (x_n^{(2)} - z_k^{(2)})^2$$

Derivative with respect to a vector

- The partial derivate of the squared difference with respect to one component

$$\frac{\partial}{\partial z_k^{(d)}} \|\boldsymbol{x}_n - \boldsymbol{z}_k\|^2 = \frac{\partial}{\partial z_k^{(d)}} \left((x_n^{(1)} - z_k^{(1)})^2 + (x_n^{(2)} - z_k^{(2)})^2 \right)$$
$$= -2(\boldsymbol{x}_n^{(d)} - \boldsymbol{z}_k^{(d)})$$

Derivative with respect to a vector

- The partial derivate of the squared difference with respect to one component

$$\frac{\partial}{\partial z_k^{(d)}} \|\boldsymbol{x}_n - \boldsymbol{z}_k\|^2 = \frac{\partial}{\partial z_k^{(d)}} \left((x_n^{(1)} - z_k^{(1)})^2 + (x_n^{(2)} - z_k^{(2)})^2 \right)$$
$$= -2(\boldsymbol{x}_n^{(d)} - \boldsymbol{z}_k^{(d)})$$

- All partial derivatives can be collected in a vector

$$\frac{\partial L}{\partial \boldsymbol{z}_k} = \begin{bmatrix} \frac{\partial L}{\partial z_k^{(1)}} \\ \frac{\partial L}{\partial z_k^{(2)}} \end{bmatrix} = \begin{bmatrix} -2(x_n^{(1)} - z_k^{(1)}) \\ -2(x_n^{(2)} - z_k^{(2)}) \end{bmatrix} = -2(\boldsymbol{x}_n - \boldsymbol{z}_k)$$

Derivative with respect to a vector

- The partial derivate of the squared difference with respect to one component

$$\frac{\partial}{\partial z_k^{(d)}} \|\boldsymbol{x}_n - \boldsymbol{z}_k\|^2 = \frac{\partial}{\partial z_k^{(d)}} \left((x_n^{(1)} - z_k^{(1)})^2 + (x_n^{(2)} - z_k^{(2)})^2 \right)$$
$$= -2(\boldsymbol{x}_n^{(d)} - \boldsymbol{z}_k^{(d)})$$

- All partial derivatives can be collected in a vector

$$\frac{\partial L}{\partial \boldsymbol{z}_k} = \begin{bmatrix} \frac{\partial L}{\partial z_k^{(1)}} \\ \frac{\partial L}{\partial z_k^{(2)}} \end{bmatrix} = \begin{bmatrix} -2(x_n^{(1)} - z_k^{(1)}) \\ -2(x_n^{(2)} - z_k^{(2)}) \end{bmatrix} = -2(\boldsymbol{x}_n - \boldsymbol{z}_k)$$

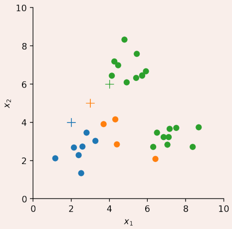
- So we have the rule

$$\frac{\partial}{\partial z_k} \|\boldsymbol{x}_n - \boldsymbol{z}_k\|^2 = -2(\boldsymbol{x}_n - \boldsymbol{z}_k)$$

Exercise: Optimal cluster center

Fix cluster assignments, optimize cluster means

$$\min_{\{\boldsymbol{z}_1,\ldots,\boldsymbol{z}_K\}} \underbrace{\sum_{k=1}^{K}}_{\text{Clusters}} \underbrace{\sum_{n:c_n=k}}_{\substack{\text{Observations}\\ \text{in cluster } k}} \|\boldsymbol{x}_n - \boldsymbol{z}_k\|^2$$

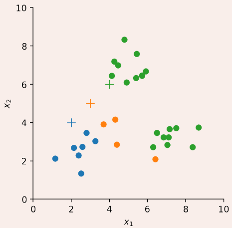


- What is the optimum value of the cluster means $\boldsymbol{z}_k$?
- Hint: Optimize the expression by computing the derivative wrt. $\boldsymbol{z}_k$, equate to zero and solve for $\boldsymbol{z}_k$

Exercise: Optimal cluster center

Fix cluster assignments, optimize cluster means

$$\min_{\{z_1,\ldots,z_K\}} \underbrace{\sum_{k=1}^{K}}_{\text{Clusters}} \underbrace{\sum_{n:c_n=k}}_{\substack{\text{Observations}\\ \text{in cluster } k}} \|\boldsymbol{x}_n - \boldsymbol{z}_k\|^2$$



- What is the optimum value of the cluster means $\boldsymbol{z}_k$?
- Hint: Optimize the expression by computing the derivative wrt. $\boldsymbol{z}_k$, equate to zero and solve for $\boldsymbol{z}_k$

Solution

$$\frac{\partial L}{\partial \boldsymbol{z}_k} \sum_{n:c_n=k} -2(\boldsymbol{x}_n - \boldsymbol{z}_k) = 2N_k\boldsymbol{z}_k - 2\sum_{n:c_n=k} \boldsymbol{x}_n = 0 \Rightarrow \boldsymbol{z}_k = \frac{1}{N_k}\sum_{n:c_n=k} \boldsymbol{x}_n$$

Exercise: Pen-and-paper k-means

Using pen-and-paper k-means, cluster the following 1-dimensional data objects

Data	$\{10, 18, 32, 70, 81, 89\}$
Num. clusters	$K = 2$
Initialization	Set means to the first two data points

Algorithm

1. Fix cluster means
   Assign each observation to closest cluster
2. Fix cluster assignments
   Set cluster means to average of data points in cluster

Exercise: K-means computational complexity

- What is the computational complexity of the k-means algorithm?
- Express it in big-O notation in terms of the number of data points N and the number of clusters K

Algorithm

1. Fix cluster means, optimize cluster assignment

$$\min_{\{c_1,\dots,c_N\}} \sum_{n=1}^{N} \|\boldsymbol{x}_n - \boldsymbol{z}_{c_n}\|^2$$

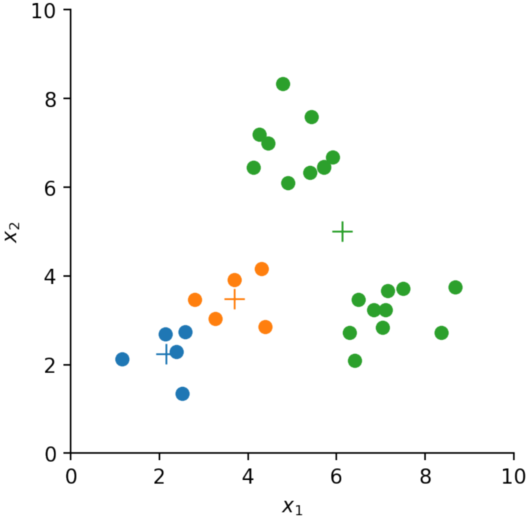
2. Fix cluster assignments, optimize cluster means

$$\min_{\{\boldsymbol{z}_1,\dots,\boldsymbol{z}_K\}} \sum_{n=1}^{N} \|\boldsymbol{x}_n - \boldsymbol{z}_{c_n}\|^2$$

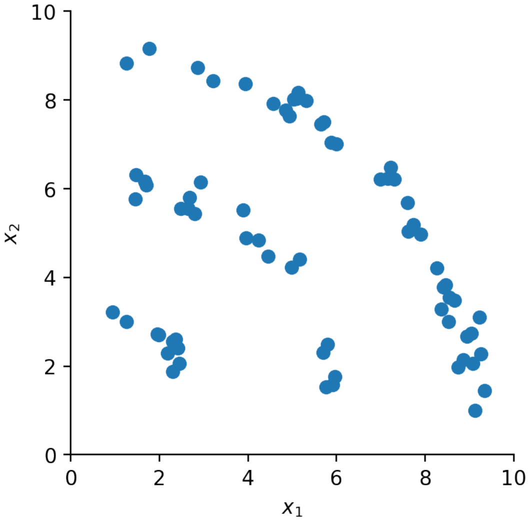
Initialization

- Algorithm requires set of initial cluster means
- Not guaranteed to converge to global optimum
- Many good heuristic initialization strategies exist

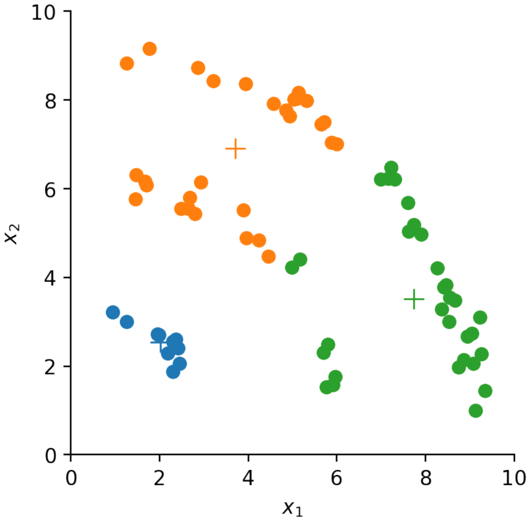
Example of local optimum



Depends on input features

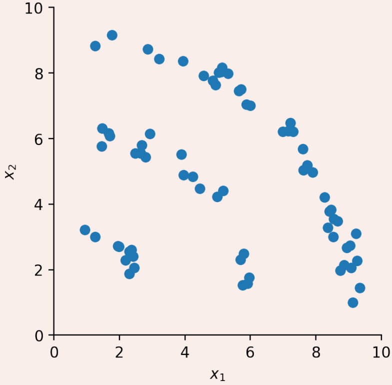


Depends on input features

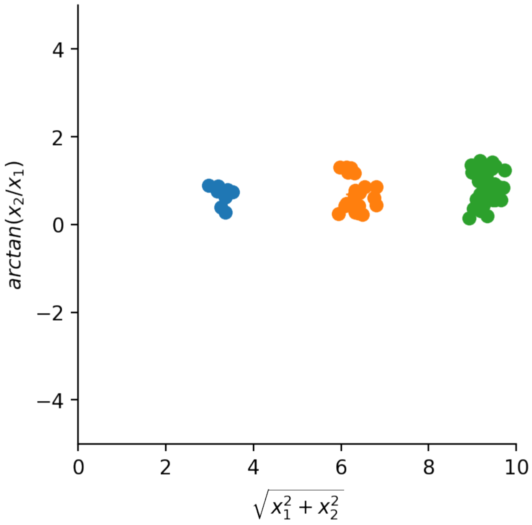


Exercise: Transformation of input features

- Can you come up with a way to transform the input features, so that k-means will find the three clusters?



Transformation of input features



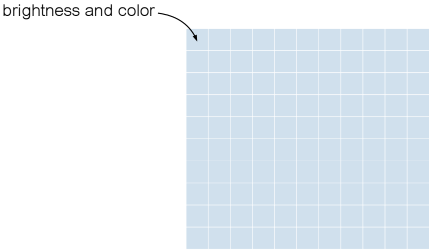
Data geometry: Image data

Exercise: What is an image?

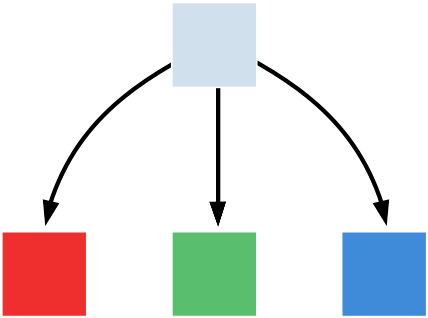
- Try to make a definition of what an *image* is without using technical terms such as pixels etc.

Image defined in technical terms

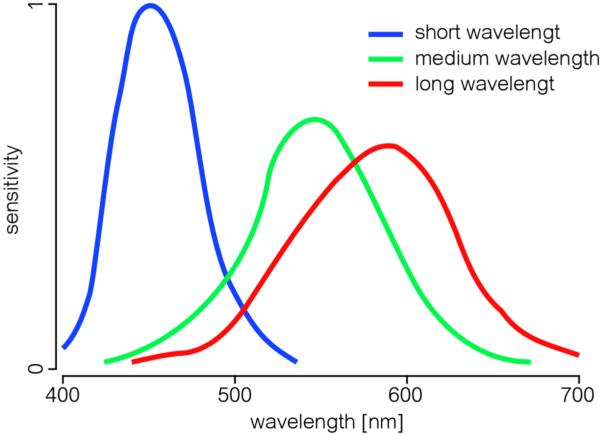
- Grid of pixels
- Every pixel contains information about color and brightness



Red, green, and blue



The eye's sensitivity to light



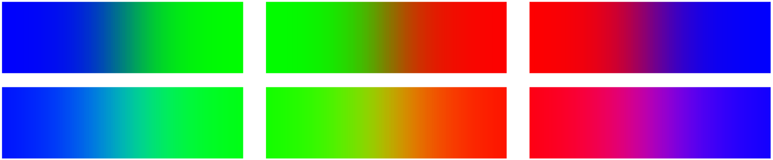
Gamma correction

- Human perception of light follows approximately a power function
- In most display systems, intensities are encoded non-linearly as

$$v_{\text{out}} = v_{\text{in}}^{\gamma}$$

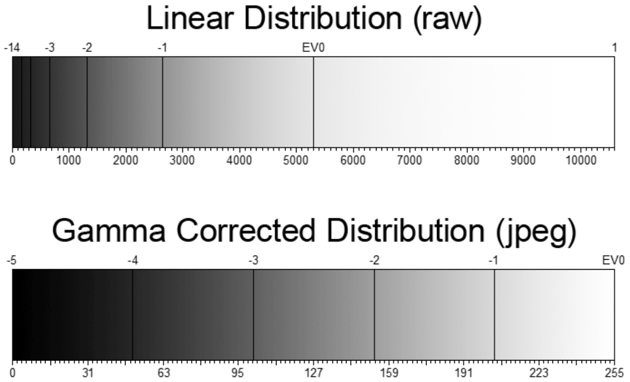
- Most systems use a value of approximately $\gamma = .45$

Example: Color gradients without and with gamma correction



Gamma correction

Example: Linear vs. gamma corrected distribution



https://en.m.wikipedia.org/wiki/File:Linear_Distribution_versus_Gamma_Corrected_Distribution.png

CIELAB color space

The CIELAB color space

- Defined by the International Commission on Illumination (CIE) in 1976
- Expresses color as three numerical values,

  L for the lightness
  A for the green-red color component
  B for the blue-yellow color components

- Perceptually uniform with respect to human color vision
- Same amount of numerical change = same amount of visually perceived change

---

https://en.wikipedia.org/wiki/CIELAB_color_space

Example: Image segmentation

Technical writing

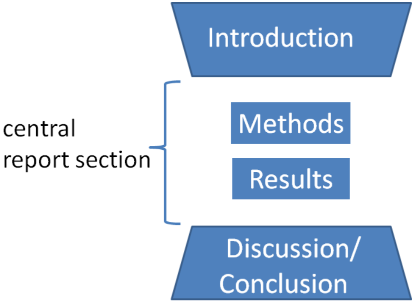
The IMRaD model

Original research articles are typically structured in this basic order

- **Introduction** Why was the study undertaken? What was the research question, the tested hypothesis or the purpose of the research?
- **Methods** When, where, and how was the study done? What materials were used or who was included in the study groups (patients, etc.)?
- **Results** What answer was found to the research question; what did the study find? Was the tested hypothesis true?
- **Discussion** What might the answer imply and why does it matter? How does it fit in with what other researchers have found? What are the perspectives for future research?

https://en.wikipedia.org/wiki/IMRAD

The IMRaD wineglass



https://en.wikipedia.org/wiki/IMRAD

IMRaD pros and cons

Pros

- Presents an idealized report of the ideas
- Clear and logical presentation
- Easy for readers to navigate

Cons

- Does not correspond to the actual research process
- Can be too rigid and simplistic

https://en.wikipedia.org/wiki/IMRAD

Abstract

- Brief summary at the beginning of a manuscript
- Help the reader ascertain the paper's purpose
- Should be a self-contained text
- Mirrors the IMRaD model

Abstract template

5-sentence abstract template

Motivation	Why do we care?
Problem	What problem will we solve? What have others done, and why is that not enough?
Approach	What is our big idea? How did we solve it? Which research, analysis, and experiments did we do?
Results	What is the answer?
Conclusions	What are the implications?

Example of abstract


```
Machine-learning tasks frequently
involve problems of manipulating
and classifying large numbers of
vectors in high-dimensional spaces.
Classical algorithms for solving
such problems typically take time
polynomial in the number of vectors
and the dimension of the space.
Quantum computers are good at
manipulating high-dimensional vectors
in large tensor product spaces.
This paper provides supervised and
unsupervised quantum machine learning
algorithms for cluster assignment
and cluster finding. Quantum machine
learning can take time logarithmic in
both the number of vectors and their
dimension, an exponential speed-up
over classical algorithms.
```


5-sentence abstract template

Motivation	Why do we care?
Problem	What problem will we solve? What have others done, and why is that not enough?
Approach	What is our big idea? How did we solve it? Which research, analysis, and experiments did we do?
Results	What is the answer?
Conclusions	What are the implications?

Lloyd, Mohsen, and Rebentrost, "Quantum algorithms for supervised and unsupervised machine learning" arXiv:1307.0411v2

Example of abstract

[Some] tasks frequently involve problems of *[something difficult]*. Classical algorithms for solving such problems typically take *[very long time]*. Quantum computers are good at *[something really complicated]*. This paper provides *[some new algorithm]*. Quantum machine learning *[is much faster than]* classical algorithms.

5-sentence abstract template

Motivation	Why do we care?
Problem	What problem will we solve? What have others done, and why is that not enough?
Approach	What is our big idea? How did we solve it? Which research, analysis, and experiments did we do?
Results	What is the answer?
Conclusions	What are the implications?

Lloyd, Mohsen, and Rebentrost, “Quantum algorithms for supervised and unsupervised machine learning” arXiv:1307.0411v2

Example of abstract

Motivation	*[Some]* tasks frequently involve problems of *[something difficult]*.
Problem	Classical algorithms for solving such problems typically take *[very long time]*.
Approach	Quantum computers are good at *[something really complicated]*.
Results	This paper provides *[some new algorithm]*.
Conclusions	Quantum machine learning *[is much faster than]* classical algorithms.

5-sentence abstract template

Motivation	Why do we care?
Problem	What problem will we solve? What have others done, and why is that not enough?
Approach	What is our big idea? How did we solve it? Which research, analysis, and experiments did we do?
Results	What is the answer?
Conclusions	What are the implications?

Lloyd, Mohsen, and Rebentrost, "Quantum algorithms for supervised and unsupervised machine learning" arXiv:1307.0411v2

Tasks

Tasks for today

Tasks today

1. Work through the two *clustering* notebooks
   `07-KMeansClustering.ipynb`
   `07-ImageSegmentation.ipynb`
2. Start working on Lab Report 3.
3. Today's feedback group
   - Carl Borg
   - Magnus Nordtorp Mabeck
   - Aleks Laith Gryn
   - Christine Amalie Meinert Cardel

Lab report hand in

- Lab 3: Image segmentation (Deadline: Thursday 26 October 20:00)