# PSY9511: Seminar 5

## Unsupervised learning

Esten H. Leonardsen

24.10.24


UNIVERSITETET
I OSLO

# Outline

1. Overview of unsupervised learning
2. Clustering
    - K-means
    - Hierarchical
3. Dimensionality reduction
    - Principal component analysis (PCA)
    - Independent component analysis (ICA)

# Unsupervised learning


UNIVERSITY
OF OSLO

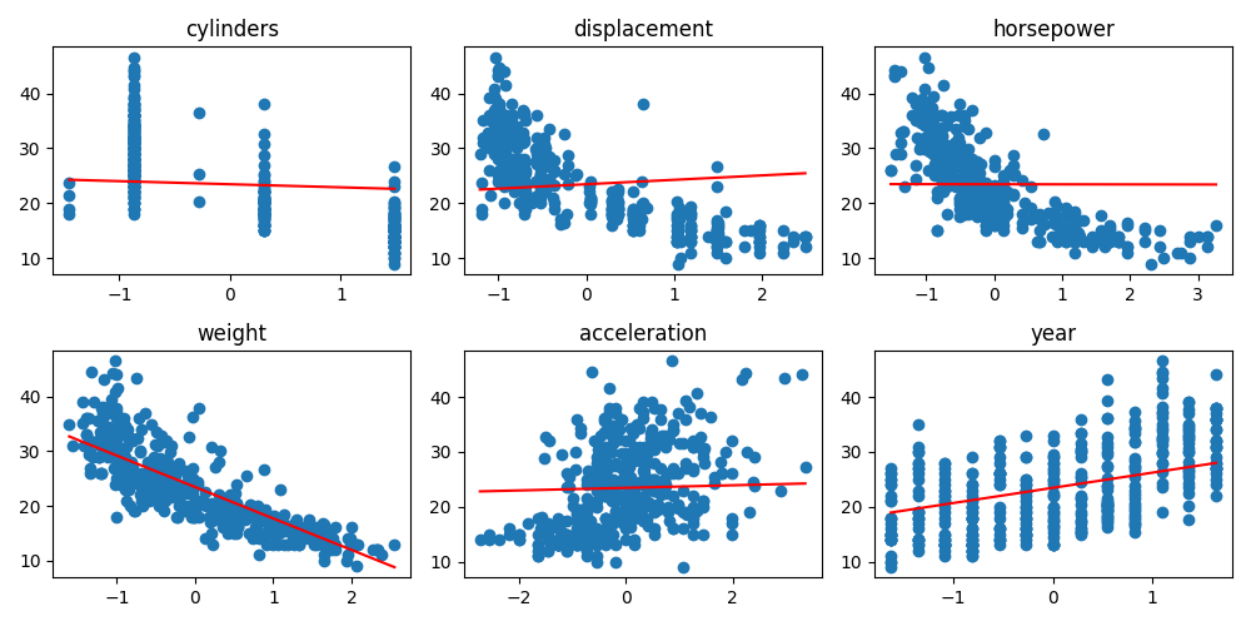
# Unsupervised learning: Motivation

Supervised learning: Find $\hat{y} = f(X)$

# Unsupervised learning: Motivation

Supervised learning: Find $\hat{y} = f(X)$

- Descriptive: Understand the relationship between $X$ and $y$



# Unsupervised learning: Motivation

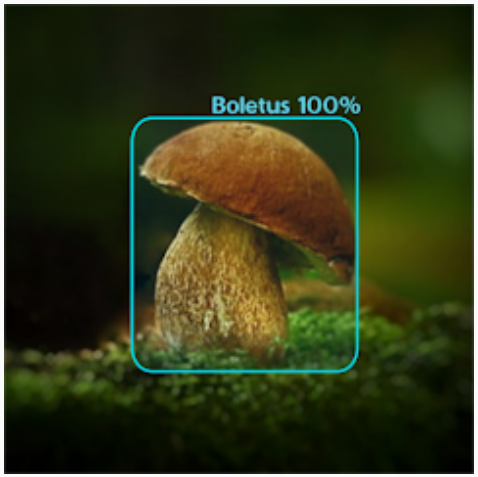
Supervised learning: Find $\hat{y} = f(X)$

- Descriptive: Understand the relationship between $X$ and $y$
- Predictive: Predict $y$ given new $X$.

# Unsupervised learning: Motivation

Supervised learning: Find $\hat{y} = f(X)$

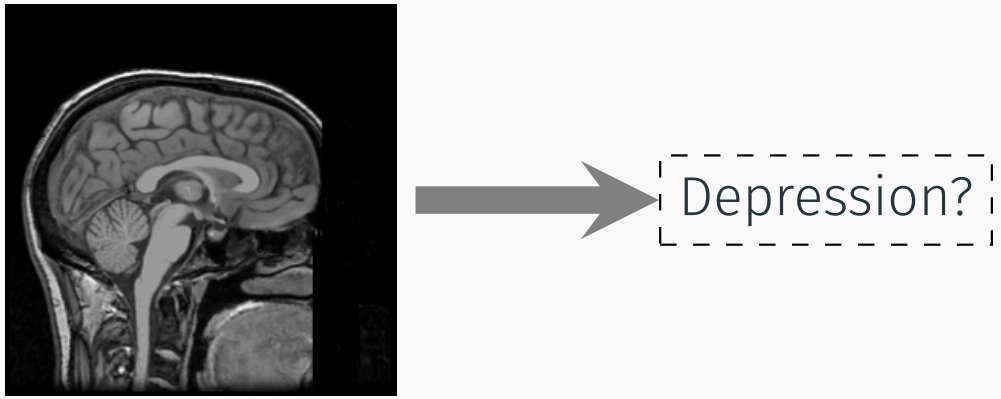
- Descriptive: Understand the relationship between $X$ and $y$
- Predictive: Predict $y$ given new $X$.
  - Because the predictions are useful



# Unsupervised learning: Motivation

**Supervised learning: Find $\hat{y} = f(X)$**

- Descriptive: Understand the relationship between $X$ and $y$
- Predictive: Predict $y$ given new $X$.
  - Because the predictions are useful
  - Because we want to know if it is possible



# Unsupervised learning: Motivation

Supervised learning: Find $\hat{y} = f(X)$

- Descriptive: Understand the relationship between $X$ and $y$
- Predictive: Predict $y$ given new $X$.
  - Because the predictions are useful
  - Because we want to know if it is possible

Unsupervised learning: Are there some interesting patterns in X?

# Unsupervised learning: Motivation

**Supervised learning: Find $\hat{y} = f(X)$**

- Descriptive: Understand the relationship between $X$ and $y$
- Predictive: Predict $y$ given new $X$.
  - Because the predictions are useful
  - Because we want to know if it is possible

**Unsupervised learning: Are there some interesting patterns in X?**

- Can we find subgroups or interesting axes of variability?

# Unsupervised learning: Motivation

**Supervised learning: Find $\hat{y} = f(X)$**

- Descriptive: Understand the relationship between $X$ and $y$
- Predictive: Predict $y$ given new $X$.
  - Because the predictions are useful
  - Because we want to know if it is possible

**Unsupervised learning: Are there some interesting patterns in X?**

- Can we find subgroups or interesting axes of variability?
- Visualization

# Clustering


UNIVERSITY
OF OSLO

# Clustering: Background

Are there some (naturally occuring) subgroups in our dataset?

# Clustering: Background

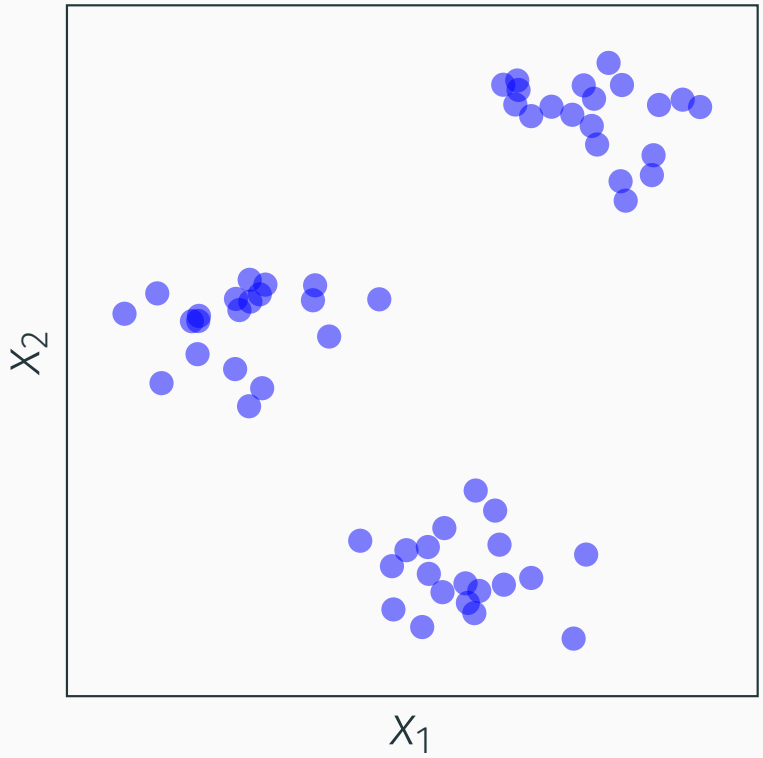
Are there some (naturally occuring) subgroups in our dataset?

| $x_1$ | $x_2$ |
|---|---|
| 0.20 | -0.26 |
| 0.15 | -0.33 |
| 0.03 | 0.07 |
| -0.07 | -0.01 |
| -0.06 | 0.00 |
| 0.28 | -0.24 |
| 0.21 | -0.35 |
| 0.20 | -0.32 |
| 0.30 | 0.25 |
| 0.00 | -0.12 |

# Clustering: Background

Are there some (naturally occuring) subgroups in our dataset?

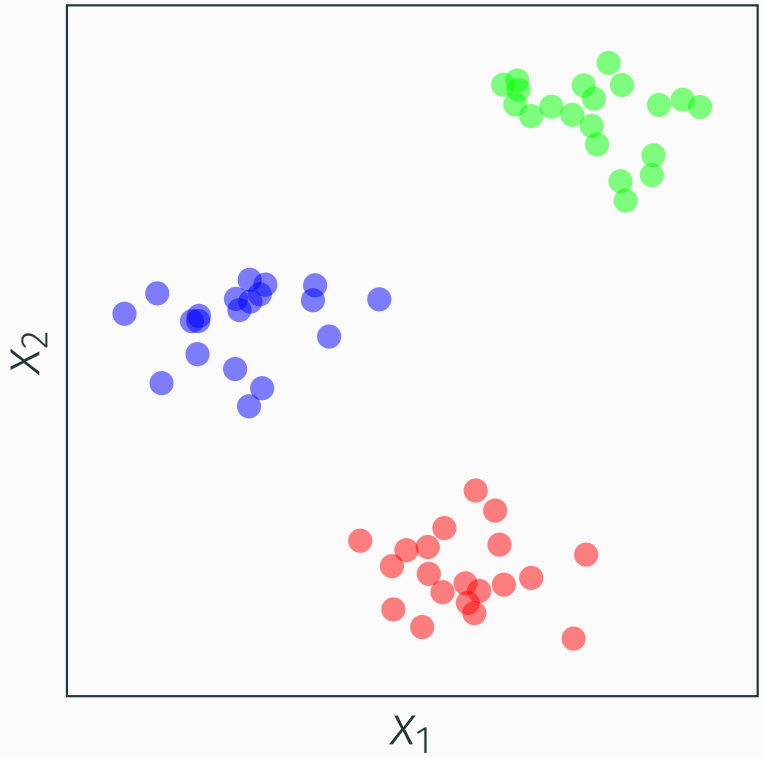
| $x_1$ | $x_2$ |
|---|---|
| 0.20 | -0.26 |
| 0.15 | -0.33 |
| 0.03 | 0.07 |
| -0.07 | -0.01 |
| -0.06 | 0.00 |
| 0.28 | -0.24 |
| 0.21 | -0.35 |
| 0.20 | -0.32 |
| 0.30 | 0.25 |
| 0.00 | -0.12 |



# Clustering: Background

Are there some (naturally occuring) subgroups in our dataset?

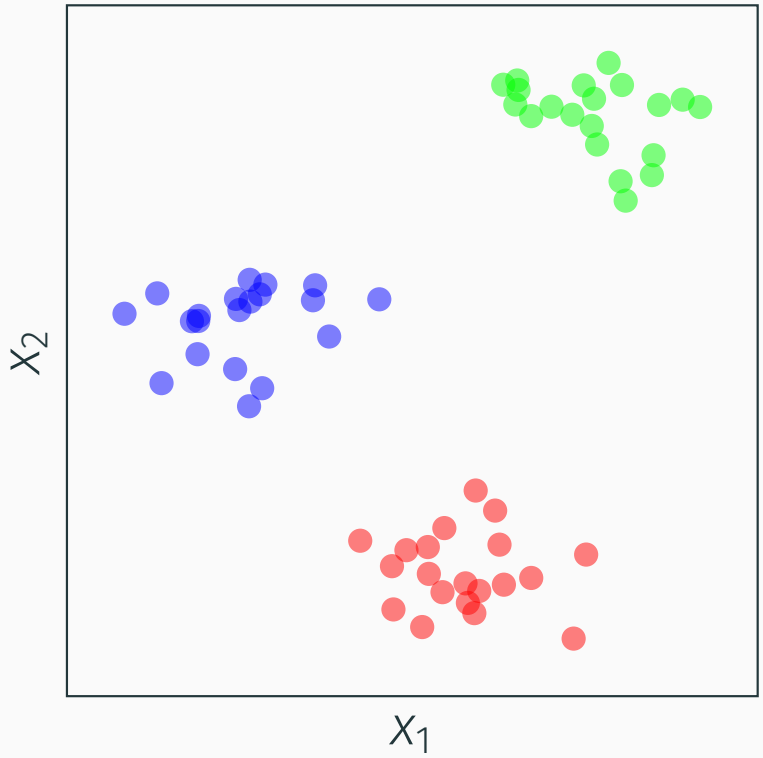
| $x_1$ | $x_2$ |
|---|---|
| 0.20 | -0.26 |
| 0.15 | -0.33 |
| 0.03 | 0.07 |
| -0.07 | -0.01 |
| -0.06 | 0.00 |
| 0.28 | -0.24 |
| 0.21 | -0.35 |
| 0.20 | -0.32 |
| 0.30 | 0.25 |
| 0.00 | -0.12 |



# Clustering: Background

Are there some (naturally occuring) subgroups in our dataset?

| $x_1$ | $x_2$ |
|---|---|
| 0.20 | -0.26 |
| 0.15 | -0.33 |
| 0.03 | 0.07 |
| -0.07 | -0.01 |
| -0.06 | 0.00 |
| 0.28 | -0.24 |
| 0.21 | -0.35 |
| 0.20 | -0.32 |
| 0.30 | 0.25 |
| 0.00 | -0.12 |



# K-means clustering

K-means clustering: Find $k$ clusters in the data to minimze the *within-cluster variance*:

$$\underset{C_1,\ldots,C_k}{\text{minimize}} \left( \sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2 \right)$$

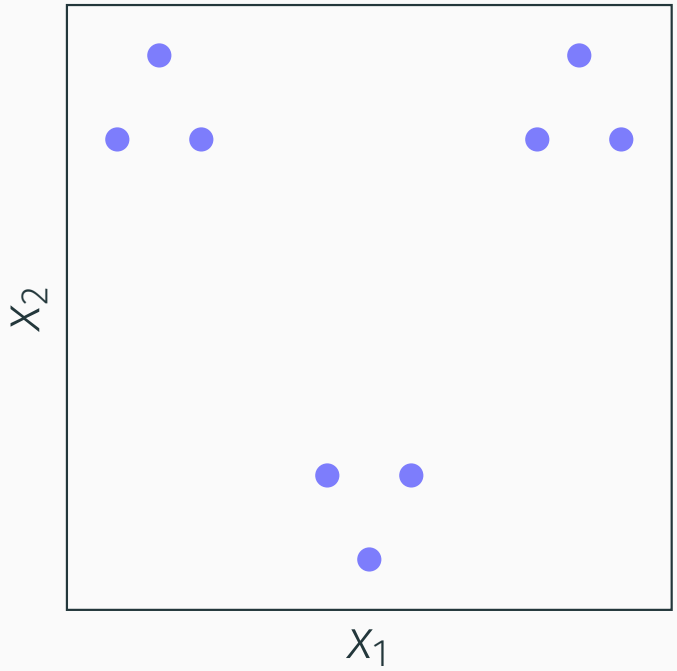
# K-means clustering

K-means clustering: Find $k$ clusters in the data to minimze the *within-cluster variance*:

$$\underset{C_1,\ldots,C_k}{\text{minimize}} \left( \sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sqrt{\sum_{j=1}^{p} (x_{ij} - x_{i'j})^2} \right)$$

# K-means clustering

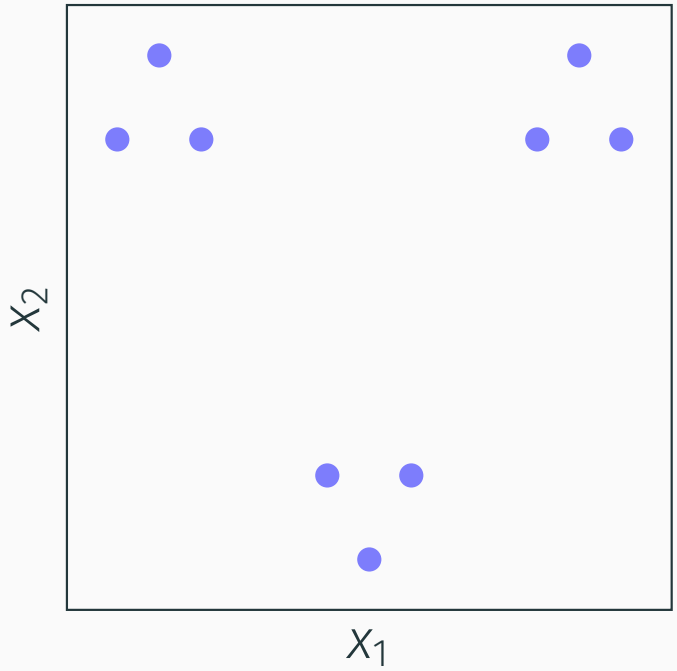
| $x_1$ | $x_2$ |
|---|---|
| 1 | 1 |
| 1.5 | 1 |
| 1.25 | 1.25 |
| -1 | 1 |
| -1.5 | 1 |
| -1.25 | 1.25 |
| 0.25 | 0 |
| -0.25 | 0 |
| 0 | -0.25 |




$$\underset{C_1,\ldots,C_k}{\text{minimize}} \left( \sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sqrt{\sum_{j=1}^{p} (x_{ij} - x_{i'j})^2} \right)$$

# K-means clustering

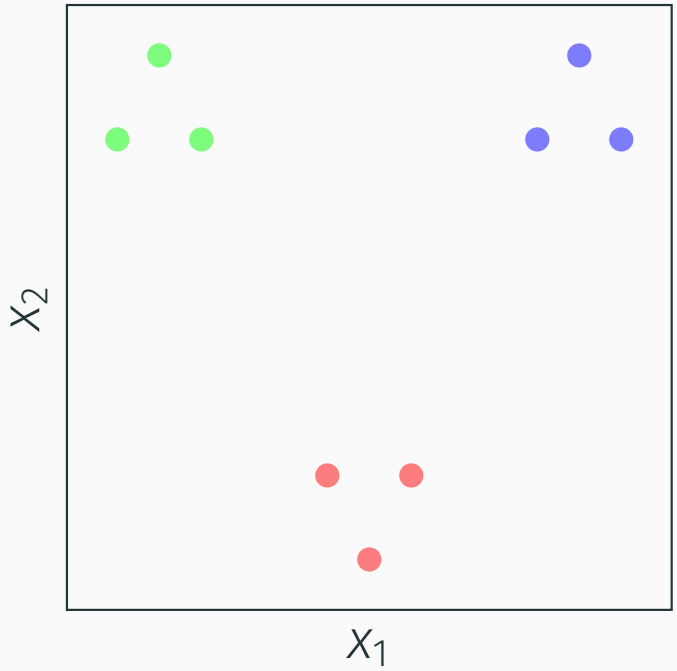
| $x_1$ | $x_2$ |
|---|---|
| 1 | 1 |
| 1.5 | 1 |
| 1.25 | 1.25 |
| -1 | 1 |
| -1.5 | 1 |
| -1.25 | 1.25 |
| 0.25 | 0 |
| -0.25 | 0 |
| 0 | -0.25 |




$$\underset{C_1,\ldots,C_k}{\text{minimize}} \left( \sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sqrt{\sum_{j=1}^{p} (x_{ij} - x_{i'j})^2} \right)$$

# K-means clustering

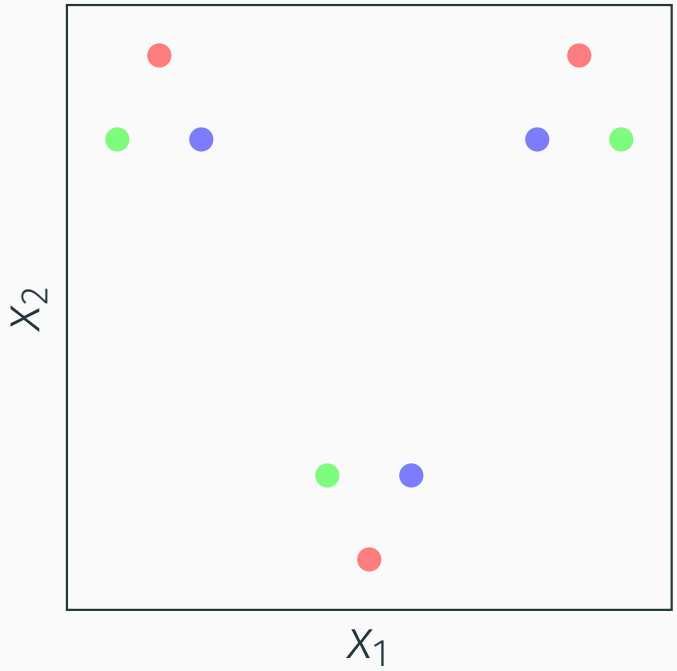
| $x_1$ | $x_2$ | |
|---|---|---|
| 1 | 1 | $C_1$ |
| 1.5 | 1 | $C_1$ |
| 1.25 | 1.25 | $C_1$ |
| -1 | 1 | $C_2$ |
| -1.5 | 1 | $C_2$ |
| -1.25 | 1.25 | $C_2$ |
| 0.25 | 0 | $C_3$ |
| -0.25 | 0 | $C_3$ |
| 0 | -0.25 | $C_3$ |




$$\underset{C_1,\ldots,C_k}{\text{minimize}} \left( \sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sqrt{\sum_{j=1}^{p} (x_{ij} - x_{i'j})^2} \right)$$

# K-means clustering

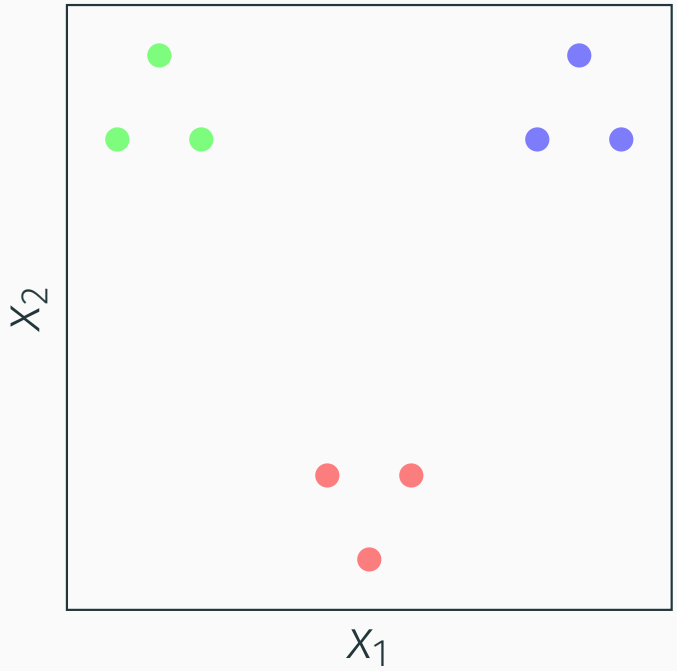
| $x_1$ | $x_2$ | |
|---|---|---|
| 1 | 1 | $C_1$ |
| 1.5 | 1 | $C_2$ |
| 1.25 | 1.25 | $C_3$ |
| -1 | 1 | $C_1$ |
| -1.5 | 1 | $C_2$ |
| -1.25 | 1.25 | $C_3$ |
| 0.25 | 0 | $C_1$ |
| -0.25 | 0 | $C_2$ |
| 0 | -0.25 | $C_3$ |




$$\underset{C_1,\ldots,C_k}{\text{minimize}} \left( \sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sqrt{\sum_{j=1}^{p} (x_{ij} - x_{i'j})^2} \right)$$

# K-means clustering

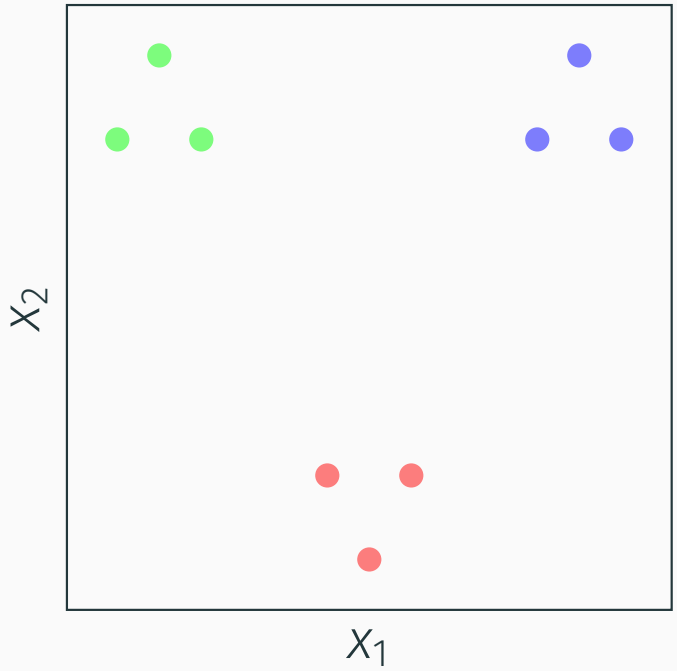
| $x_1$ | $x_2$ | |
|---|---|---|
| 1 | 1 | $C_1$ |
| 1.5 | 1 | $C_1$ |
| 1.25 | 1.25 | $C_1$ |
| -1 | 1 | $C_2$ |
| -1.5 | 1 | $C_2$ |
| -1.25 | 1.25 | $C_2$ |
| 0.25 | 0 | $C_3$ |
| -0.25 | 0 | $C_3$ |
| 0 | -0.25 | $C_3$ |




$$\underset{C_1,\ldots,C_k}{\text{minimize}} \left( \sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sqrt{\sum_{j=1}^{p} (x_{ij} - x_{i'j})^2} \right)$$

# K-means clustering

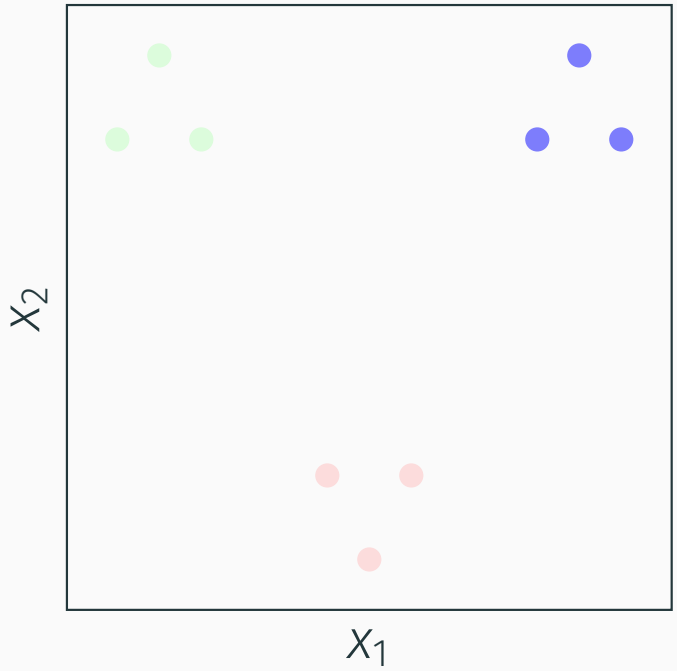
| $x_1$ | $x_2$ | |
|---|---|---|
| 1 | 1 | $C_1$ |
| 1.5 | 1 | $C_1$ |
| 1.25 | 1.25 | $C_1$ |
| -1 | 1 | $C_2$ |
| -1.5 | 1 | $C_2$ |
| -1.25 | 1.25 | $C_2$ |
| 0.25 | 0 | $C_3$ |
| -0.25 | 0 | $C_3$ |
| 0 | -0.25 | $C_3$ |




$$\underset{C_1,\ldots,C_k}{\text{minimize}} \left( \sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sqrt{\sum_{j=1}^{p} (x_{ij} - x_{i'j})^2} \right)$$

# K-means clustering

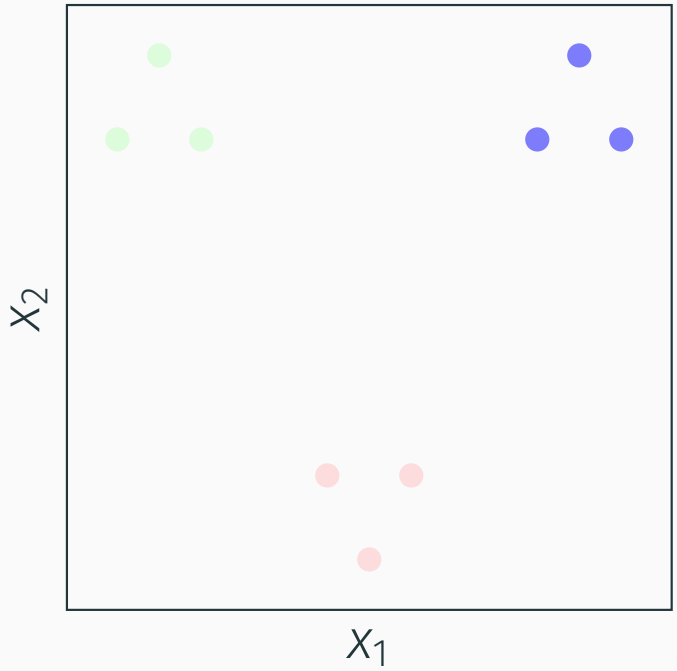
| $x_1$ | $x_2$ | |
|---|---|---|
| 1 | 1 | $C_1$ |
| 1.5 | 1 | $C_1$ |
| 1.25 | 1.25 | $C_1$ |
| -1 | 1 | $C_2$ |
| -1.5 | 1 | $C_2$ |
| -1.25 | 1.25 | $C_2$ |
| 0.25 | 0 | $C_3$ |
| -0.25 | 0 | $C_3$ |
| 0 | -0.25 | $C_3$ |




$$\underset{C_1,\ldots,C_k}{\text{minimize}} \left( \sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sqrt{\sum_{j=1}^{p} (x_{ij} - x_{i'j})^2} \right)$$

# K-means clustering

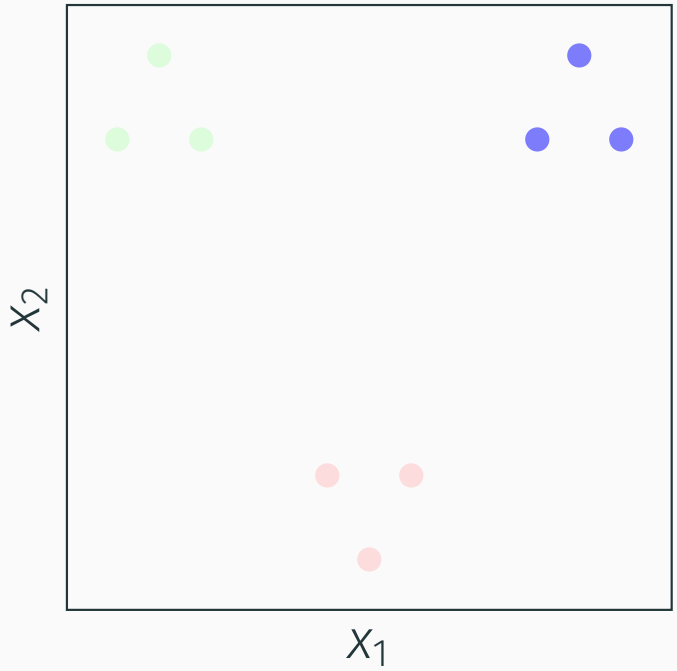
| $x_1$ | $x_2$ | |
|---|---|---|
| 1 | 1 | $C_1$ |
| 1.5 | 1 | $C_1$ |
| 1.25 | 1.25 | $C_1$ |
| -1 | 1 | $C_2$ |
| -1.5 | 1 | $C_2$ |
| -1.25 | 1.25 | $C_2$ |
| 0.25 | 0 | $C_3$ |
| -0.25 | 0 | $C_3$ |
| 0 | -0.25 | $C_3$ |




$$\underset{C_1,\ldots,C_k}{\text{minimize}} \left( \sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sqrt{\sum_{j=1}^{p} (x_{ij} - x_{i'j})^2} \right)$$

# K-means clustering

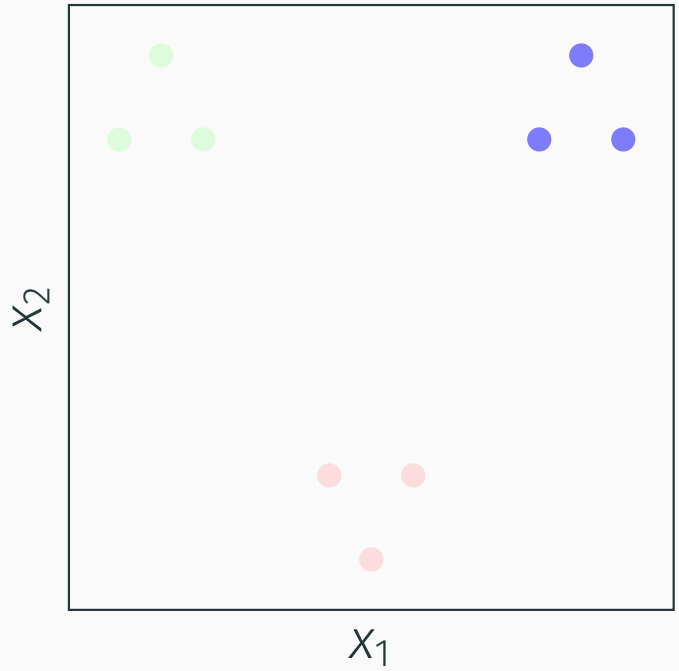
| $x_1$ | $x_2$ | |
|---|---|---|
| 1 | 1 | $C_1$ |
| 1.5 | 1 | $C_1$ |
| 1.25 | 1.25 | $C_1$ |
| -1 | 1 | $C_2$ |
| -1.5 | 1 | $C_2$ |
| -1.25 | 1.25 | $C_2$ |
| 0.25 | 0 | $C_3$ |
| -0.25 | 0 | $C_3$ |
| 0 | -0.25 | $C_3$ |




$$\underset{C_1,\ldots,C_k}{\text{minimize}} \left( \sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sqrt{\sum_{j=1}^{p} (x_{ij} - x_{i'j})^2} \right)$$

# K-means clustering

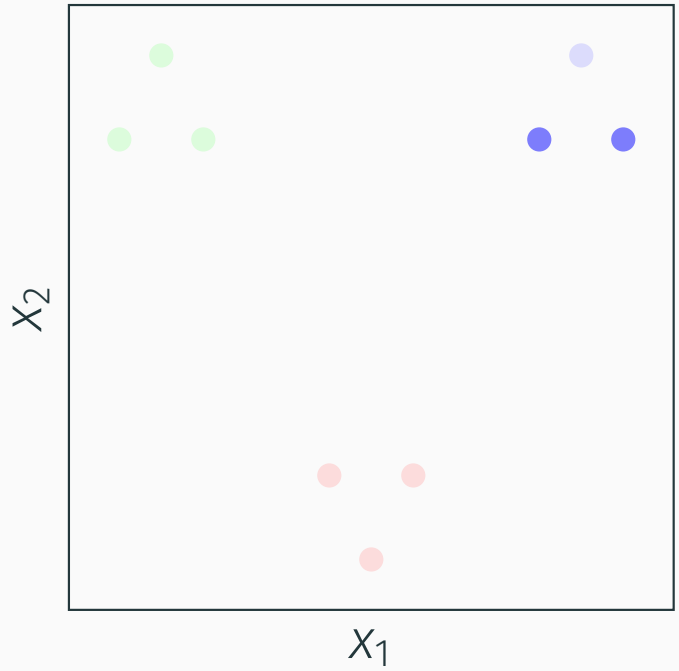
| $i$ | $x_1$ | $x_2$ | |
|---|---|---|---|
| | 1 | 1 | $C_1$ |
| | 1.5 | 1 | $C_1$ |
| | 1.25 | 1.25 | $C_1$ |
| | -1 | 1 | $C_2$ |
| | -1.5 | 1 | $C_2$ |
| | -1.25 | 1.25 | $C_2$ |
| | 0.25 | 0 | $C_3$ |
| | -0.25 | 0 | $C_3$ |
| | 0 | -0.25 | $C_3$ |




$$\underset{C_1,\ldots,C_k}{\text{minimize}} \left( \sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sqrt{\sum_{j=1}^{p} (x_{ij} - x_{i'j})^2} \right)$$

# K-means clustering

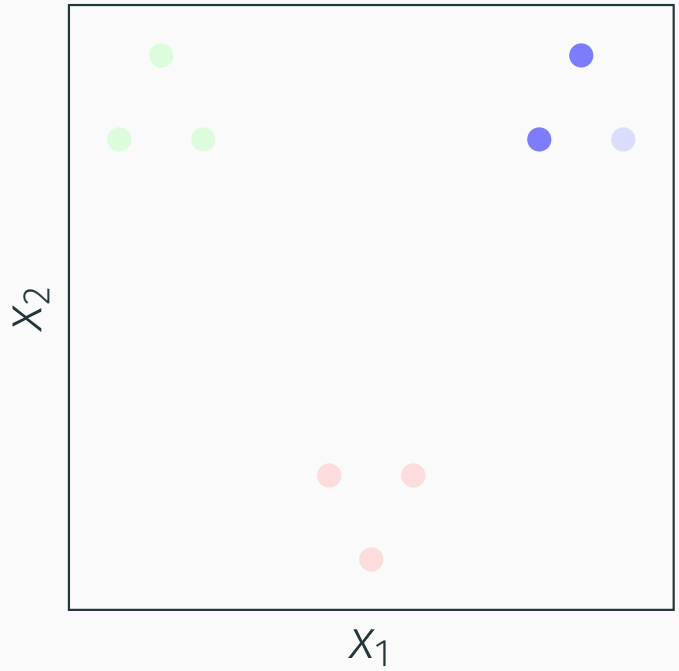
| $x_1$ | $x_2$ |
|---|---|
| 1 | 1 |
| 1.5 | 1 |
| 1.25 | 1.25 |
| -1 | 1 |
| -1.5 | 1 |
| -1.25 | 1.25 |
| 0.25 | 0 |
| -0.25 | 0 |
| 0 | -0.25 |




$$\underset{C_1,\ldots,C_k}{\text{minimize}} \left( \sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sqrt{\sum_{j=1}^{p} (x_{ij} - x_{i'j})^2} \right)$$

# K-means clustering

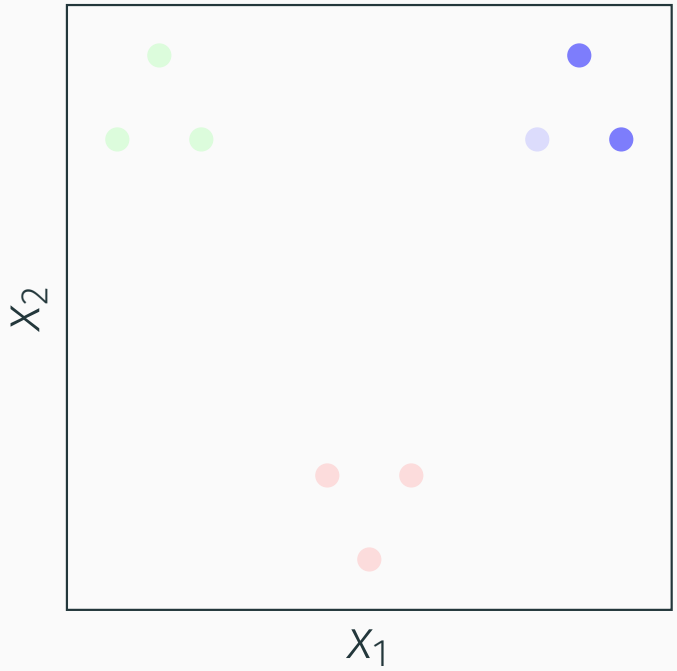
| $x_1$ | $x_2$ |
|---|---|
| 1 | 1 |
| 1.5 | 1 |
| 1.25 | 1.25 |
| -1 | 1 |
| -1.5 | 1 |
| -1.25 | 1.25 |
| 0.25 | 0 |
| -0.25 | 0 |
| 0 | -0.25 |




$$\underset{C_1,\ldots,C_k}{\text{minimize}} \left( \sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sqrt{\sum_{j=1}^{p} (x_{ij} - x_{i'j})^2} \right)$$

# K-means clustering

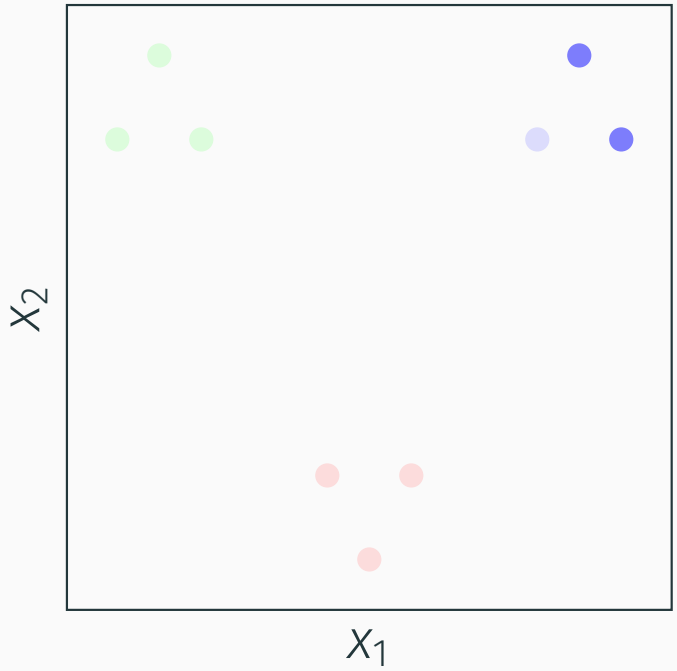
| $x_1$ | $x_2$ |
|---|---|
| 1 | 1 |
| 1.5 | 1 |
| 1.25 | 1.25 |
| -1 | 1 |
| -1.5 | 1 |
| -1.25 | 1.25 |
| 0.25 | 0 |
| -0.25 | 0 |
| 0 | -0.25 |




$$\underset{C_1,\ldots,C_k}{\text{minimize}} \left( \sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sqrt{\sum_{j=1}^{p} (x_{ij} - x_{i'j})^2} \right)$$

# K-means clustering

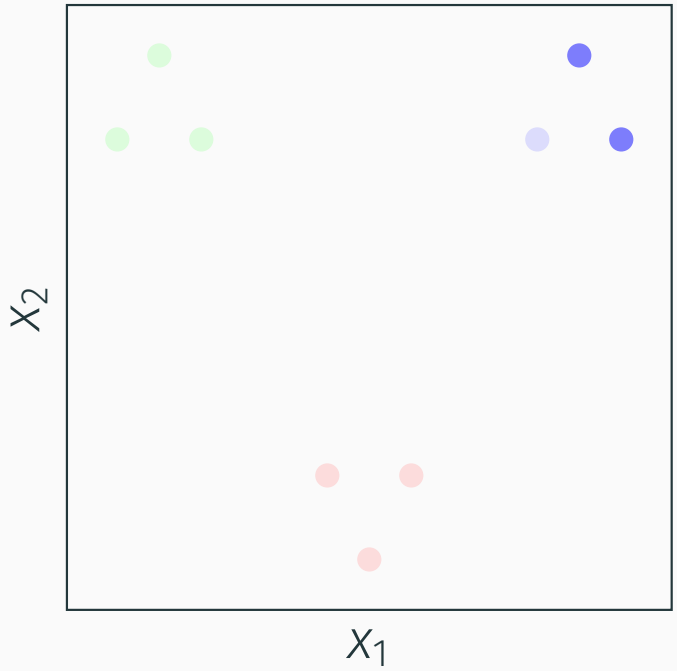
| $x_1$ | $x_2$ |
|---|---|
| 1 | 1 |
| 1.5 | 1 |
| 1.25 | 1.25 |
| -1 | 1 |
| -1.5 | 1 |
| -1.25 | 1.25 |
| 0.25 | 0 |
| -0.25 | 0 |
| 0 | -0.25 |




$$\underset{C_1,\ldots,C_k}{\text{minimize}} \left( \sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sqrt{\sum_{j=1}^{p} (x_{ij} - x_{i'j})^2} \right)$$

# K-means clustering

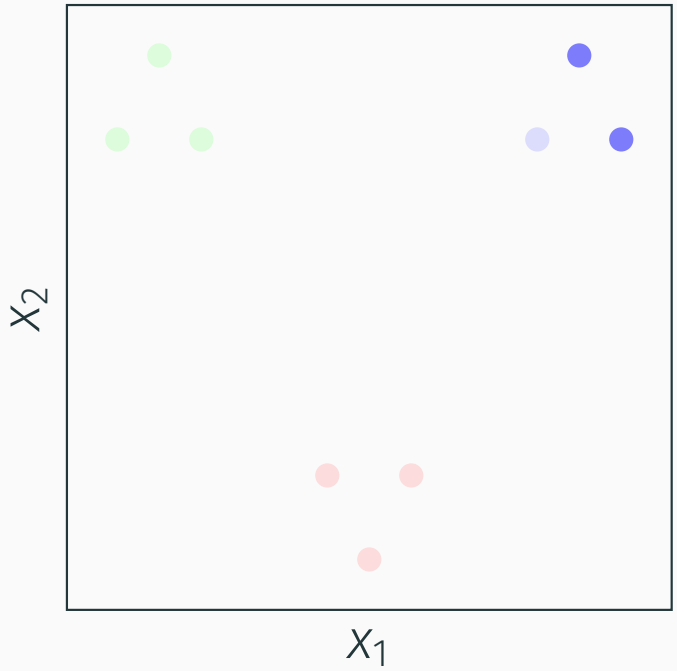
| $x_1$ | $x_2$ |
|---|---|
| 1 | 1 |
| 1.5 | 1 |
| 1.25 | 1.25 |
| -1 | 1 |
| -1.5 | 1 |
| -1.25 | 1.25 |
| 0.25 | 0 |
| -0.25 | 0 |
| 0 | -0.25 |




$$\underset{C_1,\ldots,C_k}{\text{minimize}} \left( \sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sqrt{\sum_{j=1}^{p} (x_{ij} - x_{i'j})^2} \right)$$

# K-means clustering


| $x_1$ | $x_2$ |
|---|---|
| 1 | 1 |
| 1.5 | 1 |
| 1.25 | 1.25 |
| -1 | 1 |
| -1.5 | 1 |
| -1.25 | 1.25 |
| 0.25 | 0 |
| -0.25 | 0 |
| 0 | -0.25 |




$$\underset{C_1,\ldots,C_k}{\text{minimize}} \left( \sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sqrt{\sum_{j=1}^{p} (x_{ij} - x_{i'j})^2} \right)$$

# K-means clustering

| $x_1$ | $x_2$ |
|---|---|
| 1 | 1 |
| 1.5 | 1 |
| 1.25 | 1.25 |
| -1 | 1 |
| -1.5 | 1 |
| -1.25 | 1.25 |
| 0.25 | 0 |
| -0.25 | 0 |
| 0 | -0.25 |

$$\underset{C_1,\ldots,C_k}{\text{minimize}} \left( \sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sqrt{\sum_{j=1}^{p} (x_{ij} - x_{i'j})^2} \right)$$

# K-means clustering

| $x_1$ | $x_2$ |
|---|---|
| 1 | 1 |
| 1.5 | 1 |
| 1.25 | 1.25 |
| -1 | 1 |
| -1.5 | 1 |
| -1.25 | 1.25 |
| 0.25 | 0 |
| -0.25 | 0 |
| 0 | -0.25 |

$$\underset{C_1,\ldots,C_k}{\text{minimize}} \left( \sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sqrt{\sum_{j=1}^{p} (x_{ij} - x_{i'j})^2} \right)$$