PSY9511: Seminar 5

Unsupervised learning

Esten H. Leonardsen

24.10.24

UNIVERSITY OF OSLO

Outline

1. Exercise 4
2. Overview of unsupervised learning
3. Clustering
    - K-means
    - Hierarchical
4. Dimensionality reduction
    - Principal component analysis (PCA)
    - Independent component analysis (ICA)
    - Partial least squares (PLS)

Exercise 4

Exercise 4: Stratification

http://localhost:8888/notebooks/notebooks%2FStratification.ipynb

Exercise 4: Solution

http://localhost:8888/notebooks/notebooks/Solution%204.ipynb

Unsupervised learning

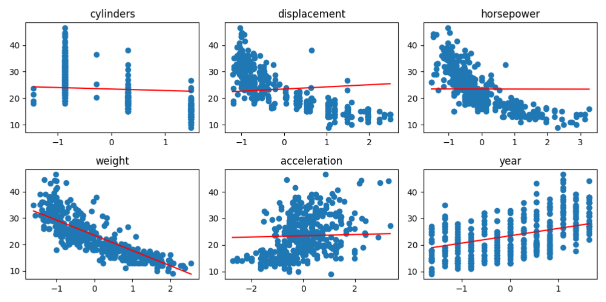
UNIVERSITY
OF OSLO

Unsupervised learning: Motivation

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

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

- Descriptive: Understand the relationship between X and y



Unsupervised learning: Motivation

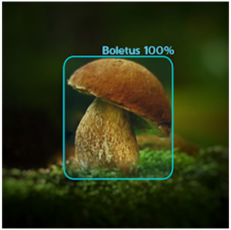
Supervised learning: Find $\hat{y} = f(X)$

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

Unsupervised learning: Motivation

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

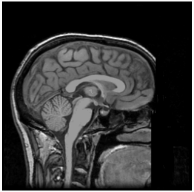
- Descriptive: Understand the relationship between X and y
- Predictive: Predict y given new X.
  - Because the predictions are useful



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

 → Depression?

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 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 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 interesting patterns in X?

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

Clustering

Clustering: Background

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

Clustering: Background

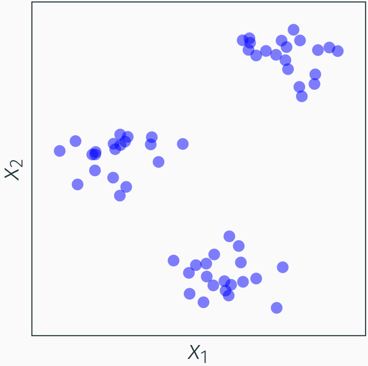
Are there some (naturally occuring) subgroups in our dataset?

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?

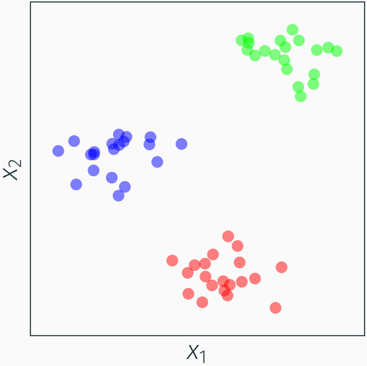
x_1	x_2
0.20	-0.26
0.15	-0.33
0.03	0.07
-0.07	-0.01
-0.06	0.00
0.28	-0.24
0.21	-0.35
0.20	-0.32
0.30	0.25
0.00	-0.12



Clustering: Background

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

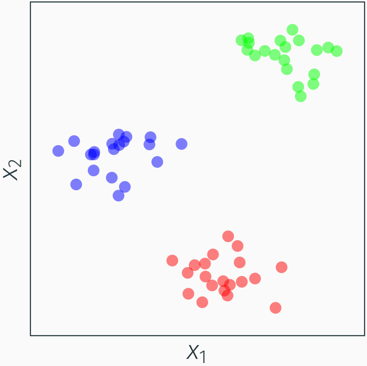
x_1	x_2
0.20	-0.26
0.15	-0.33
0.03	0.07
-0.07	-0.01
-0.06	0.00
0.28	-0.24
0.21	-0.35
0.20	-0.32
0.30	0.25
0.00	-0.12



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: Definition

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

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

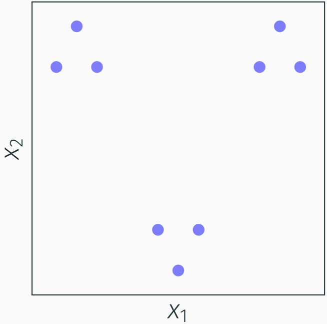
K-means clustering: Definition

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: Definition

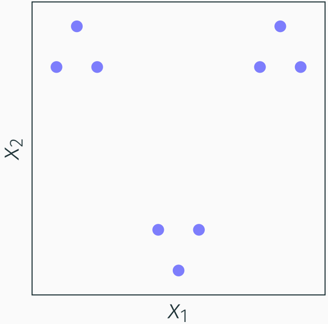
x_1	x_2
1	1
1.5	1
1.25	1.25
-1	1
-1.5	1
-1.25	1.25
0.25	0
-0.25	0
0	-0.25



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

K-means clustering: Definition

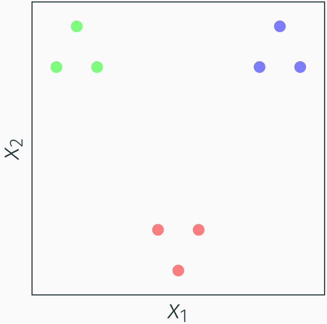
x_1	x_2
1	1
1.5	1
1.25	1.25
-1	1
-1.5	1
-1.25	1.25
0.25	0
-0.25	0
0	-0.25



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

K-means clustering: Definition

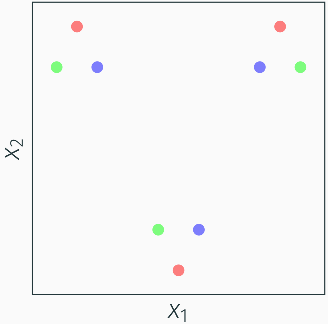
x_1	x_2	
1	1	C_1
1.5	1	C_1
1.25	1.25	C_1
-1	1	C_2
-1.5	1	C_2
-1.25	1.25	C_2
0.25	0	C_3
-0.25	0	C_3
0	-0.25	C_3



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

K-means clustering: Definition

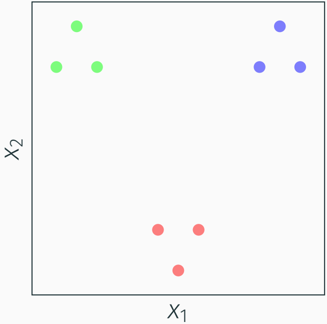
x_1	x_2	
1	1	C_1
1.5	1	C_2
1.25	1.25	C_3
-1	1	C_1
-1.5	1	C_2
-1.25	1.25	C_3
0.25	0	C_1
-0.25	0	C_2
0	-0.25	C_3



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

K-means clustering: Definition

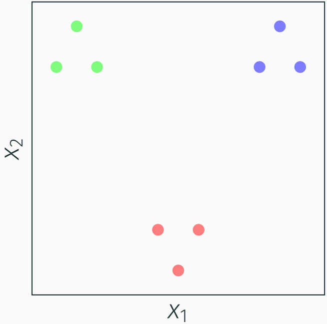
x_1	x_2	
1	1	C_1
1.5	1	C_1
1.25	1.25	C_1
-1	1	C_2
-1.5	1	C_2
-1.25	1.25	C_2
0.25	0	C_3
-0.25	0	C_3
0	-0.25	C_3



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

K-means clustering: Definition

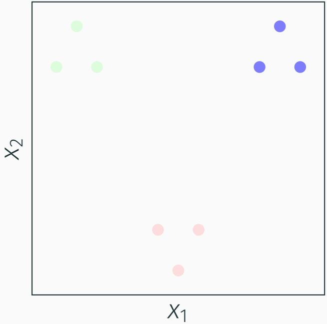
x_1	x_2	
1	1	C_1
1.5	1	C_1
1.25	1.25	C_1
-1	1	C_2
-1.5	1	C_2
-1.25	1.25	C_2
0.25	0	C_3
-0.25	0	C_3
0	-0.25	C_3



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

K-means clustering: Definition

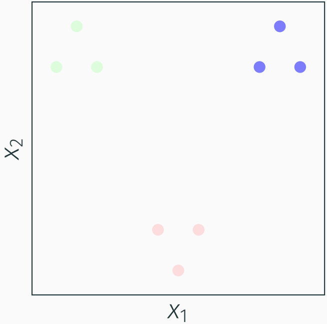
x_1	x_2	
1	1	C_1
1.5	1	C_1
1.25	1.25	C_1
-1	1	C_2
-1.5	1	C_2
-1.25	1.25	C_2
0.25	0	C_3
-0.25	0	C_3
0	-0.25	C_3



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

K-means clustering: Definition

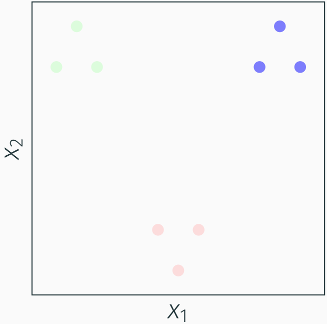
x_1	x_2	
1	1	C_1
1.5	1	C_1
1.25	1.25	C_1
-1	1	C_2
-1.5	1	C_2
-1.25	1.25	C_2
0.25	0	C_3
-0.25	0	C_3
0	-0.25	C_3



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

K-means clustering: Definition

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

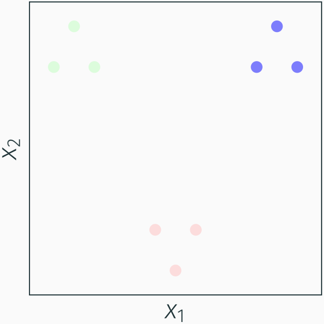


$$\underset{C_1,\ldots,C_k}{\text{minimize}} \left(\sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sqrt{\sum_{j=1}^{p} (x_{ij} - x_{i'j})^2} \right)$$

K-means clustering: Definition

i, i'

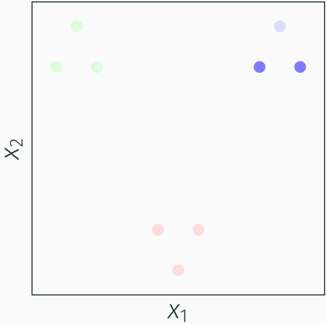
x_1	x_2	
1	1	C_1
1.5	1	C_1
1.25	1.25	C_1
-1	1	C_2
-1.5	1	C_2
-1.25	1.25	C_2
0.25	0	C_3
-0.25	0	C_3
0	-0.25	C_3



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

K-means clustering: Definition

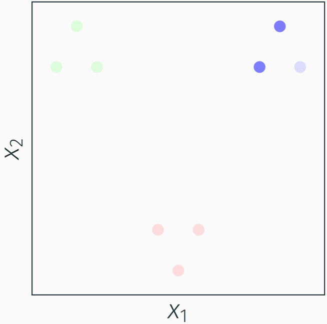
x_1	x_2
1	1
1.5	1
1.25	1.25
-1	1
-1.5	1
-1.25	1.25
0.25	0
-0.25	0
0	-0.25



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

K-means clustering: Definition

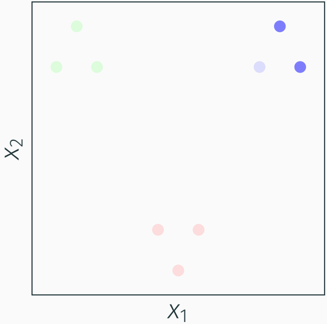
x_1	x_2
1	1
1.5	1
1.25	1.25
-1	1
-1.5	1
-1.25	1.25
0.25	0
-0.25	0
0	-0.25



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

K-means clustering: Definition

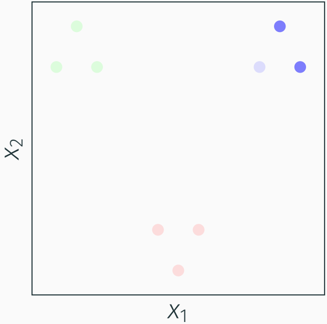
x_1	x_2
1	1
1.5	1
1.25	1.25
-1	1
-1.5	1
-1.25	1.25
0.25	0
-0.25	0
0	-0.25



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

K-means clustering: Definition

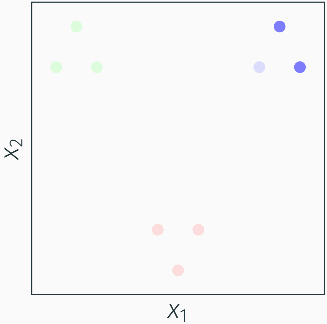
x_1	x_2
1	1
1.5	1
1.25	1.25
-1	1
-1.5	1
-1.25	1.25
0.25	0
-0.25	0
0	-0.25



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

K-means clustering: Definition

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

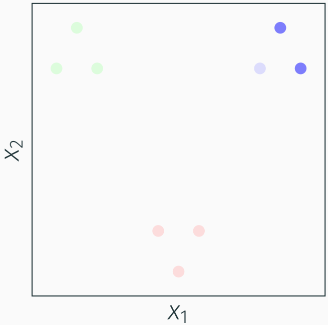


$$\underset{C_1,\ldots,C_k}{\text{minimize}} \left(\sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sqrt{\sum_{j=1}^{p} (x_{ij} - x_{i'j})^2} \right)$$

K-means clustering: Definition

← j →

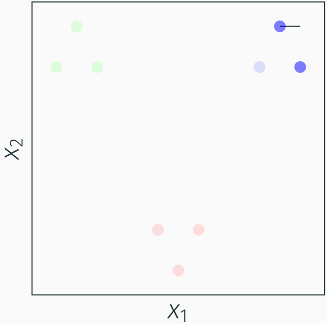
x_1	x_2
1	1
1.5	1
1.25	1.25
-1	1
-1.5	1
-1.25	1.25
0.25	0
-0.25	0
0	-0.25



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

K-means clustering: Definition

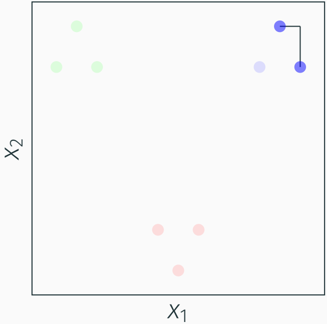
x_1	x_2
1	1
1.5	1
1.25	1.25
-1	1
-1.5	1
-1.25	1.25
0.25	0
-0.25	0
0	-0.25



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

K-means clustering: Definition

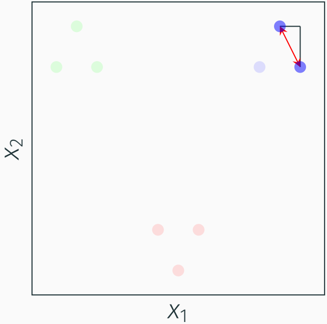
x_1	x_2
1	1
1.5	1
1.25	1.25
-1	1
-1.5	1
-1.25	1.25
0.25	0
-0.25	0
0	-0.25



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

K-means clustering: Definition

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



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

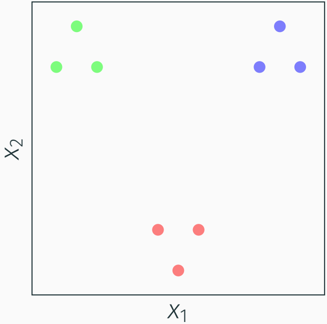
K-means clustering: Definition

We want to find an assignment of clusters that for each cluster, for each pair of points within the cluster, minimizes the distance between them

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

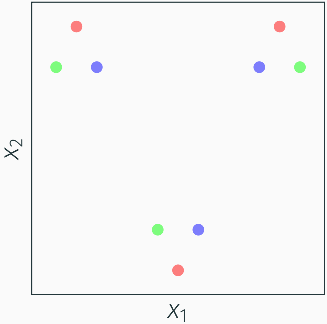
K-means clustering: Definition

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



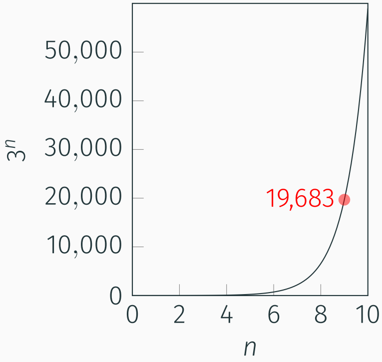
K-means clustering: Definition

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



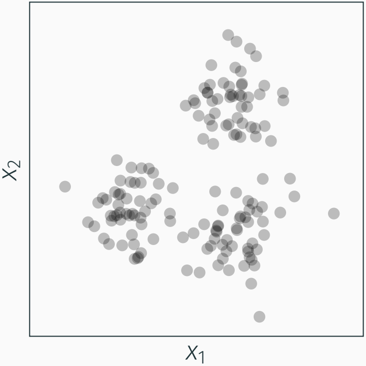
K-means clustering: Definition

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



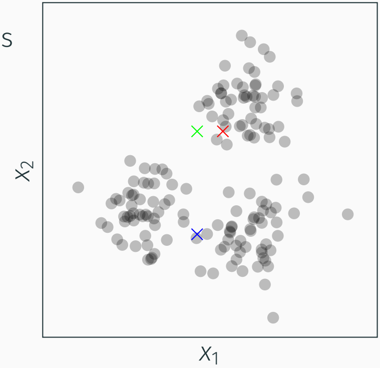
Number of possible assignments: K^n

K-means clustering: Algorithm



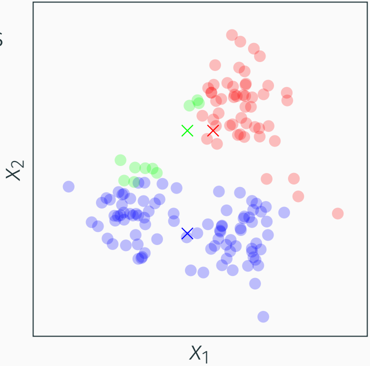
K-means clustering: Algorithm

1. Initialize k random centroids



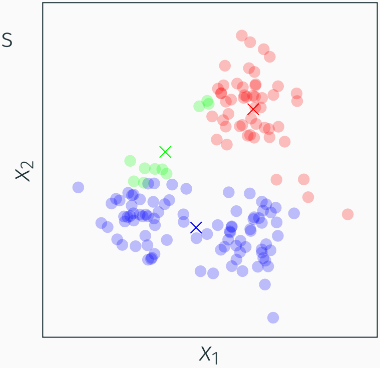
K-means clustering: Algorithm

1. Initialize k random centroids
2. Iteratively:
   2.1 Assign each point to the nearest centroid



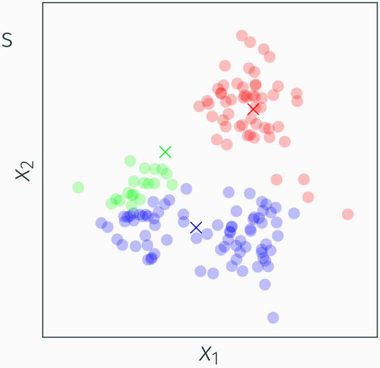
K-means clustering: Algorithm

1. Initialize k random centroids
2. Iteratively:
   2.1 Assign each point to the nearest centroid
   2.2 Update the centroids



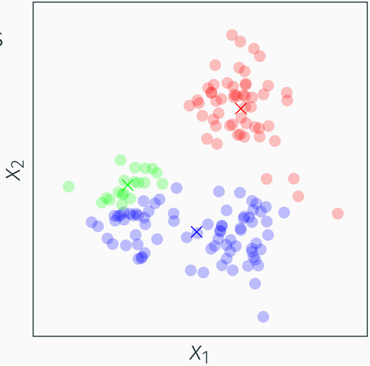
K-means clustering: Algorithm

1. Initialize k random centroids
2. Iteratively:
   2.1 Assign each point to the nearest centroid
   2.2 Update the centroids



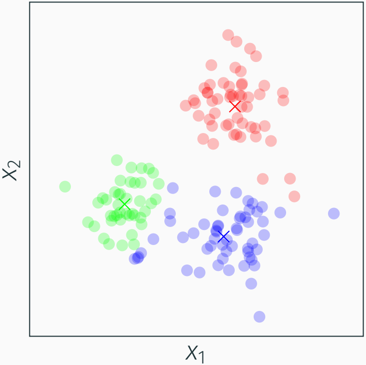
K-means clustering: Algorithm

1. Initialize k random centroids
2. Iteratively:
   2.1 Assign each point to the nearest centroid
   2.2 Update the centroids



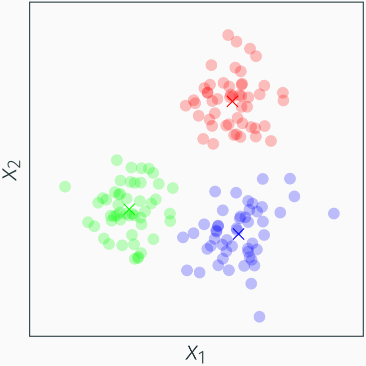
K-means clustering: Algorithm

1. Initialize k random centroids
2. Iteratively:
   2.1 Assign each point to the nearest centroid
   2.2 Update the centroids



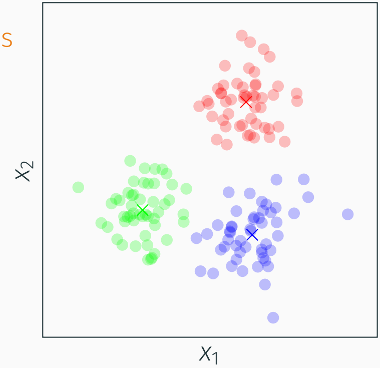
K-means clustering: Algorithm

1. Initialize k random centroids
2. Iteratively:
   2.1 Assign each point to the nearest centroid
   2.2 Update the centroids



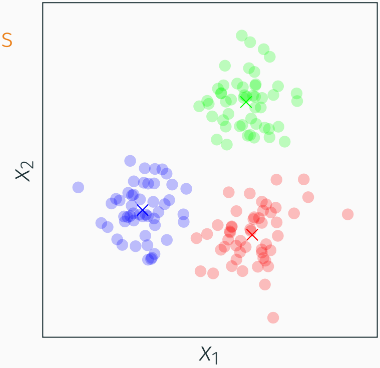
K-means clustering: Algorithm

1. Initialize k random centroids
2. Iteratively:
   2.1 Assign each point to the nearest centroid
   2.2 Update the centroids



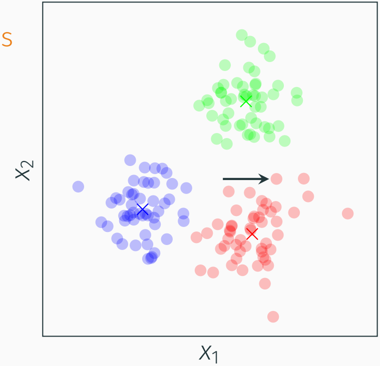
K-means clustering: Algorithm

1. Initialize k random centroids
2. Iteratively:
   2.1 Assign each point to the nearest centroid
   2.2 Update the centroids



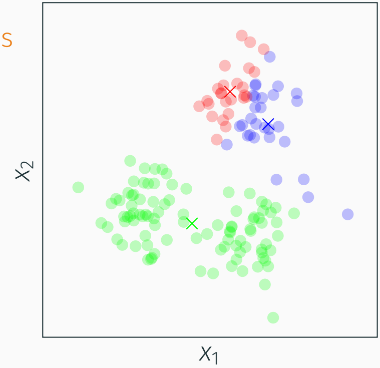
K-means clustering: Algorithm

1. Initialize k random centroids
2. Iteratively:
    2.1 Assign each point to the nearest centroid
    2.2 Update the centroids

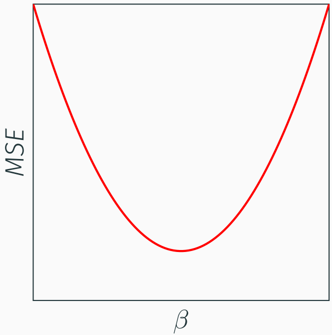


K-means clustering: Algorithm

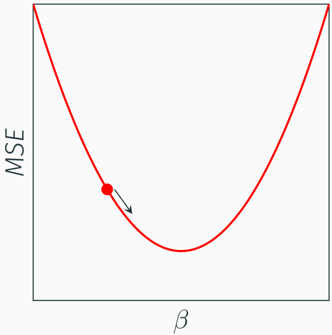
1. Initialize *k* random centroids
2. Iteratively:
   2.1 Assign each point to the nearest centroid
   2.2 Update the centroids



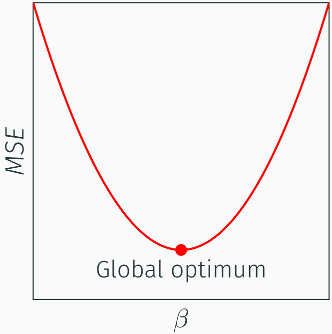
K-means clustering: Algorithm



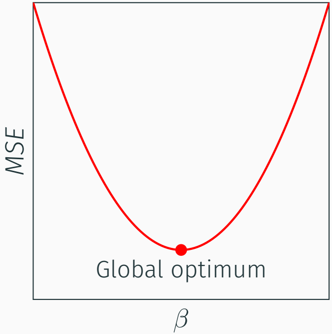
K-means clustering: Algorithm

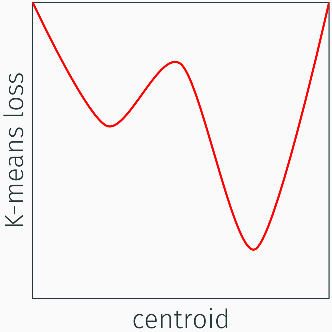


K-means clustering: Algorithm

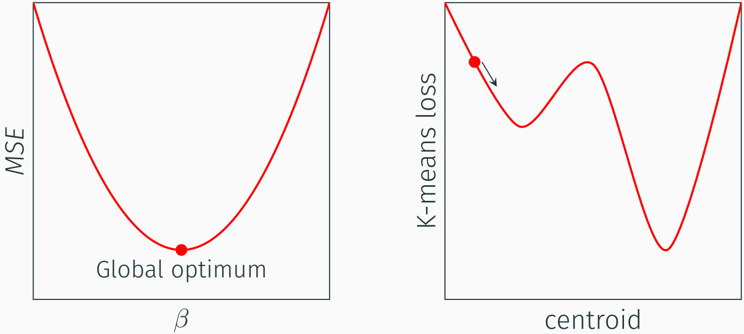


K-means clustering: Algorithm

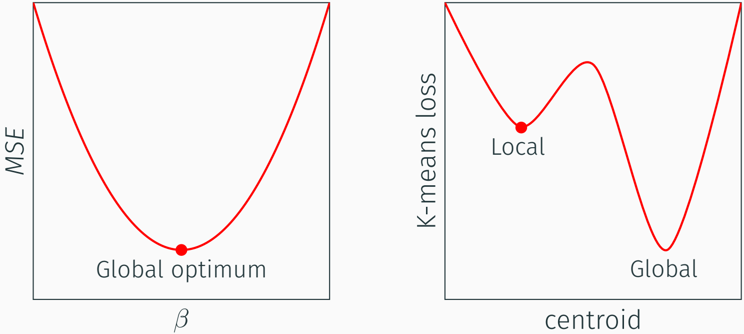




K-means clustering: Algorithm

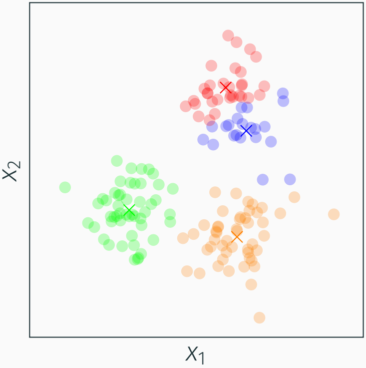


K-means clustering: Algorithm



K-means clustering: Algorithm

1. Initialize k random centroids
2. Iteratively:
   2.1 Assign each point to the nearest centroid
   2.2 Update the centroids

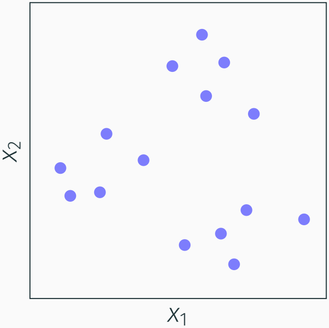


K-means clustering: Summary

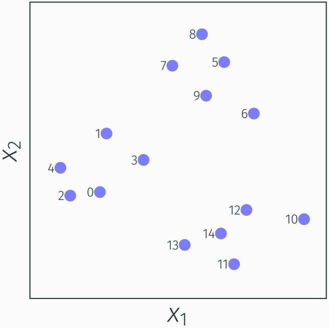
K-means clustering: Finds clusters by minimizing the within-cluster variance

- + Intuitive way of finding clusters
- + Fast algorithm
- – Dependent on knowing the number of clusters to use
- – Dependent on the random initialization of centroids

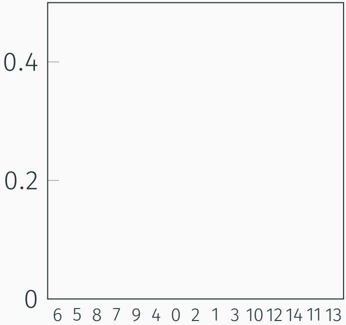
Hierarchical clustering: Algorithm

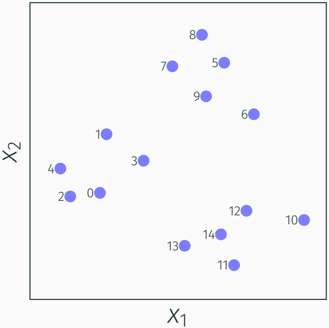


Hierarchical clustering: Algorithm

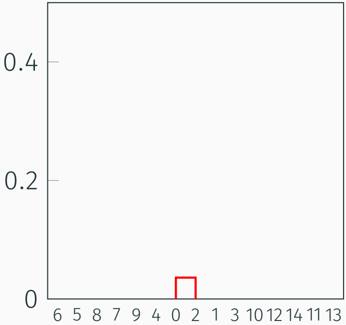


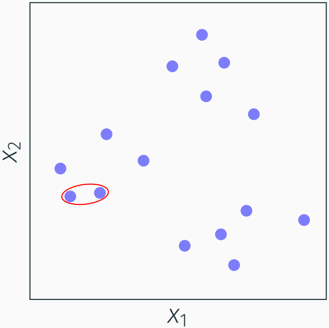
Hierarchical clustering: Algorithm



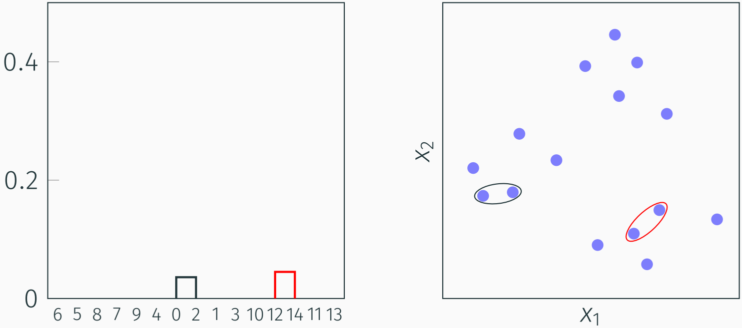


Hierarchical clustering: Algorithm

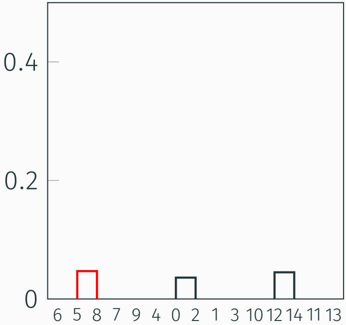


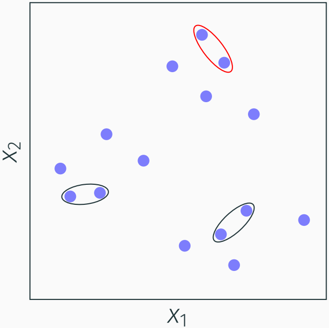


Hierarchical clustering: Algorithm

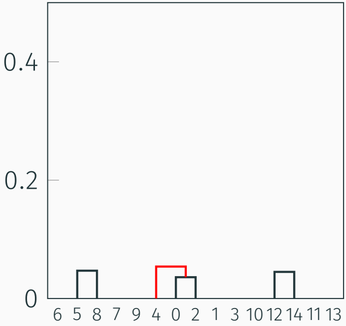


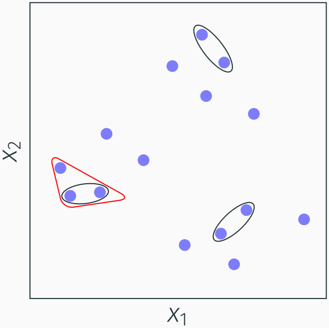
Hierarchical clustering: Algorithm



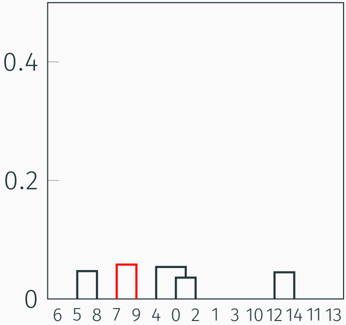


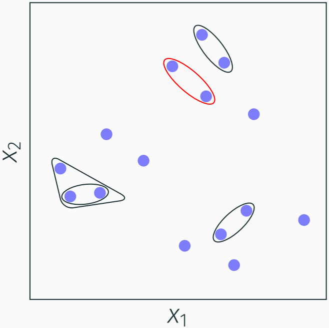
Hierarchical clustering: Algorithm



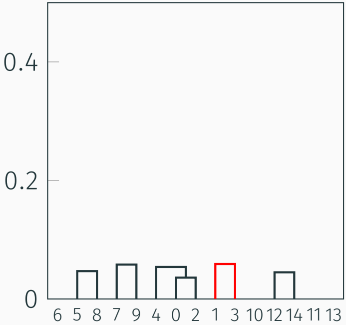


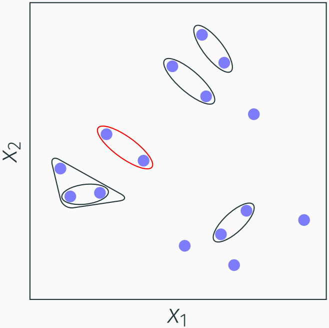
Hierarchical clustering: Algorithm



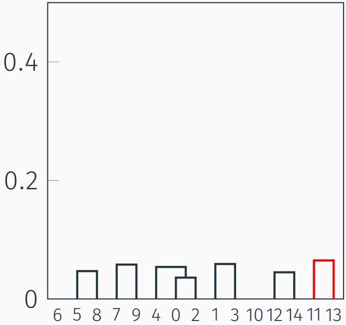


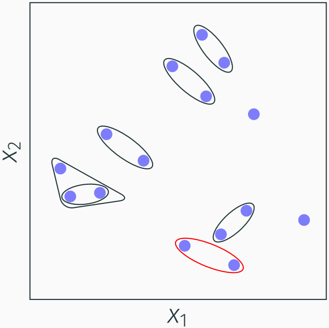
Hierarchical clustering: Algorithm

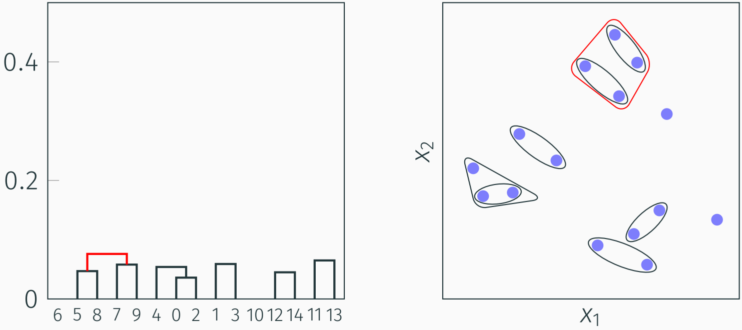




Hierarchical clustering: Algorithm

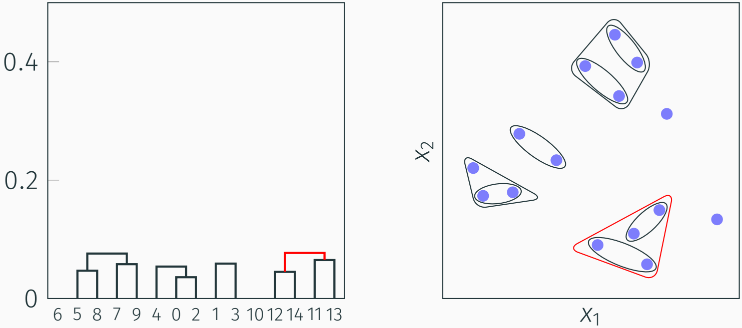




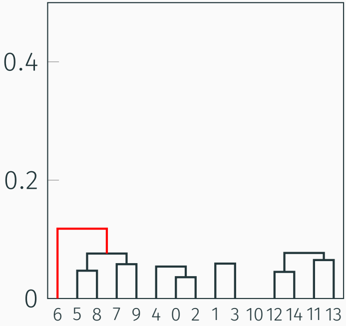


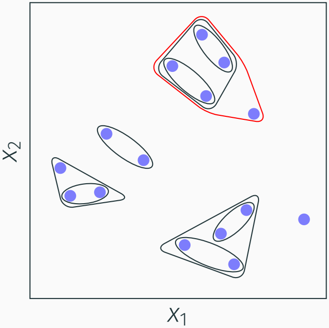
0.4
0.2
0
6 5 8 7 9 4 0 2 1 3 10 12 14 11 13
X_2
X_1

Hierarchical clustering: Algorithm

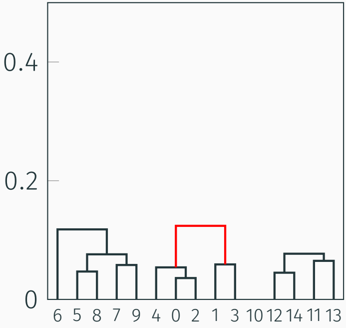


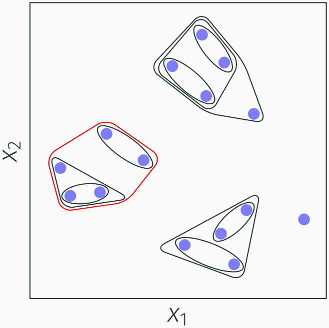
Hierarchical clustering: Algorithm



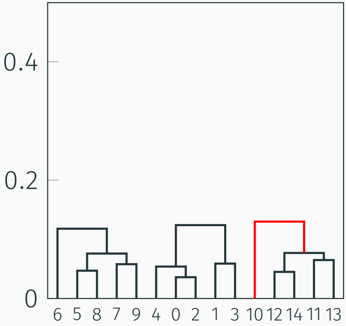


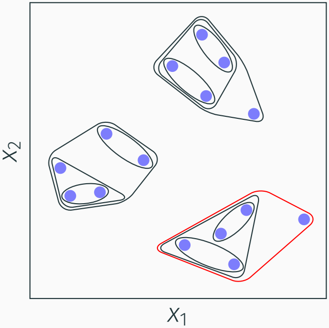
Hierarchical clustering: Algorithm



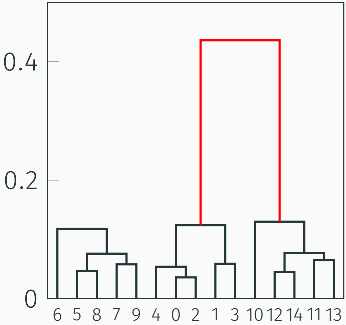


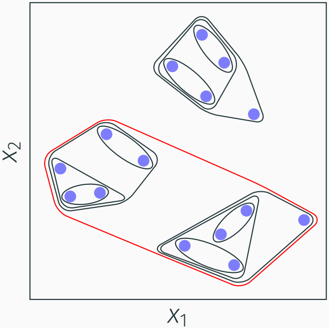
Hierarchical clustering: Algorithm



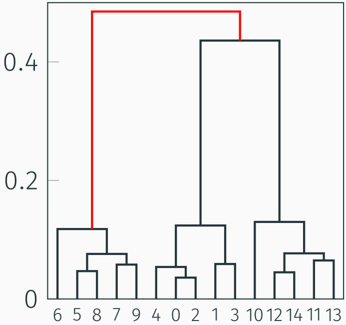


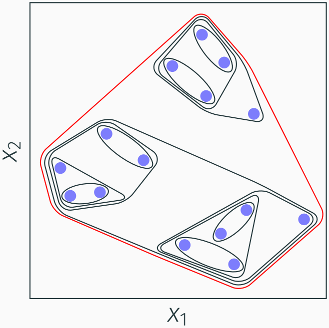
Hierarchical clustering: Algorithm



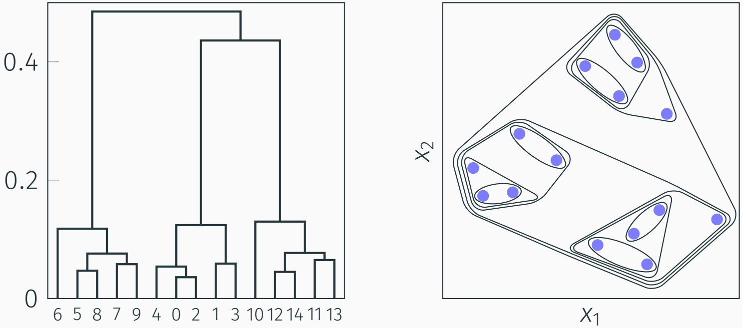


Hierarchical clustering: Algorithm

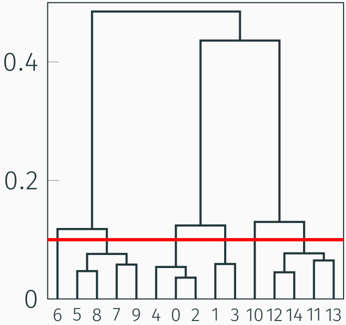


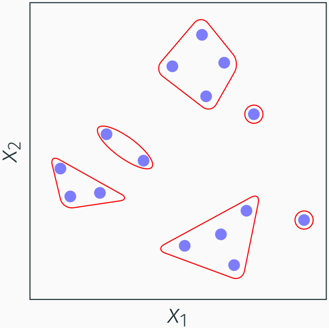


Hierarchical clustering: Algorithm

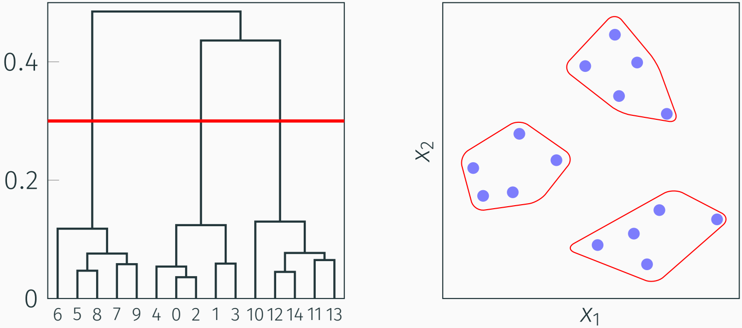


Hierarchical clustering: Interpretation

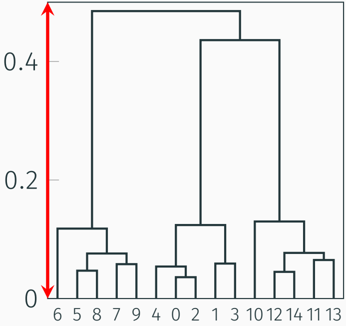


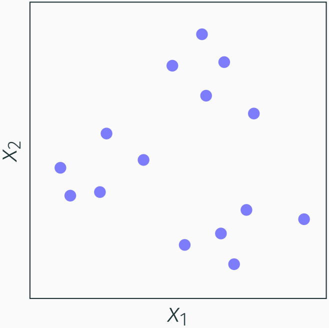


Hierarchical clustering: Interpretation

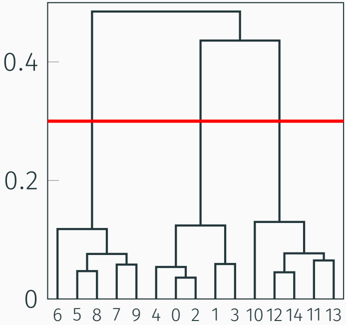


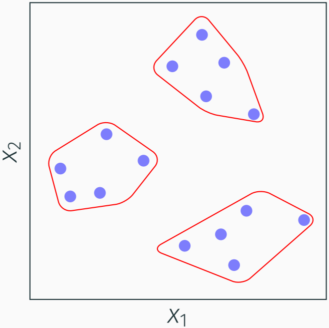
Hierarchical clustering: Interpretation



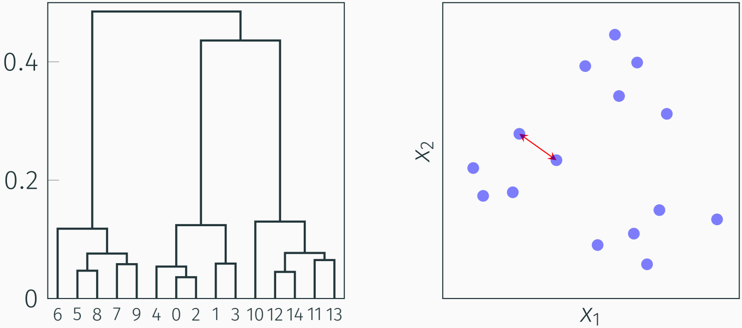


Hierarchical clustering: Interpretation

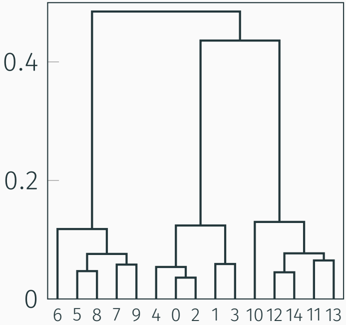


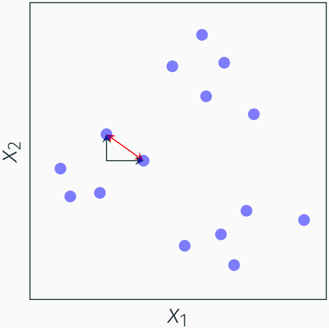


Hierarchical clustering: Interpretation

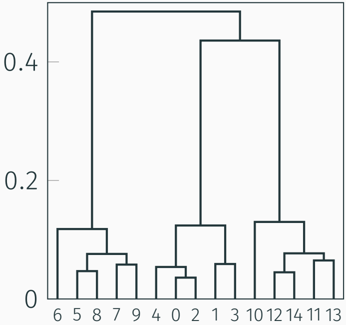


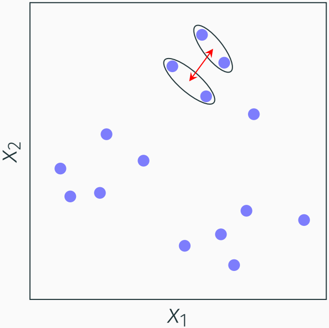
Hierarchical clustering: Interpretation



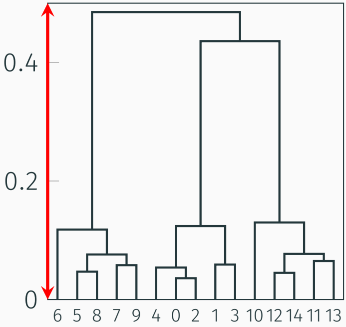


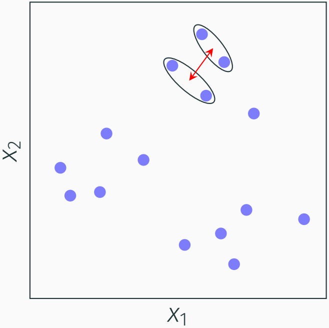
Hierarchical clustering: Interpretation





Hierarchical clustering: Interpretation





Hierarchical clustering: Interpretation

Complete	Maximal intercluster dissimilarity
Single	Minimal intercluster dissimilarity
Average	Mean intercluster dissimilarity
Centroid	Dissimilarity between the centroid for cluster A and the centroid for cluster B.

Hierarchical clustering: Interpretation

Agglomerative clustering: Iteratively merge clusters to form a hierarchy of cluster assignments.

+ Not reliant on *a priori* deciding the number of clusters
- Still relies on choices: Distance metric, linkage method, threshold

☠️☠️☠️ Clustering horror story ☠️☠️☠️

UNIVERSITY OF OSLO

☠️☠️☠️ Clustering horror story ☠️☠️☠️

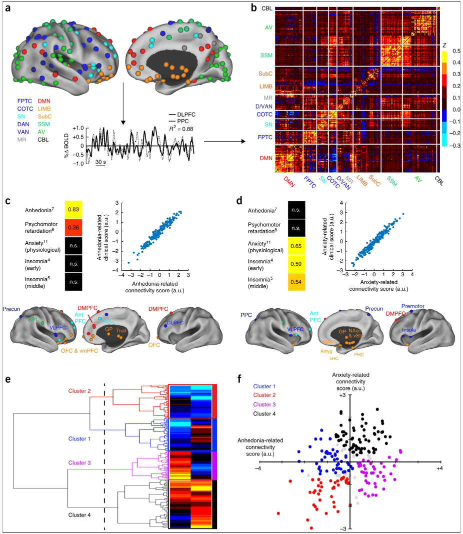
Resting-state connectivity biomarkers define neurophysiological subtypes of depression

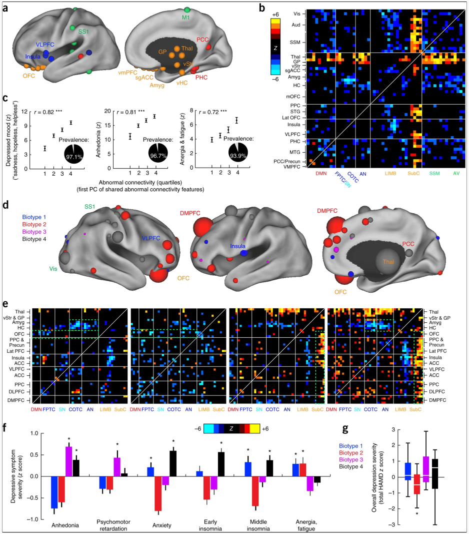
Andrew T Drysdale, Logan Grosenick, Jonathan Downar, Katharine Dunlop, Farrokh Mansouri, Yue Meng, Robert N Fetcho, Benjamin Zebley, Desmond J Oathes, Amit Etkin, Alan F Schatzberg, Keith Sudheimer, Jennifer Keller, Helen S Mayberg, Faith M Gunning, George S Alexopoulos, Michael D Fox, Alvaro Pascual-Leone, Henning U Voss, BJ Casey, Marc J Dubin & Conor Liston

Nature Medicine **23**, 28–38 (2017) | Cite this article

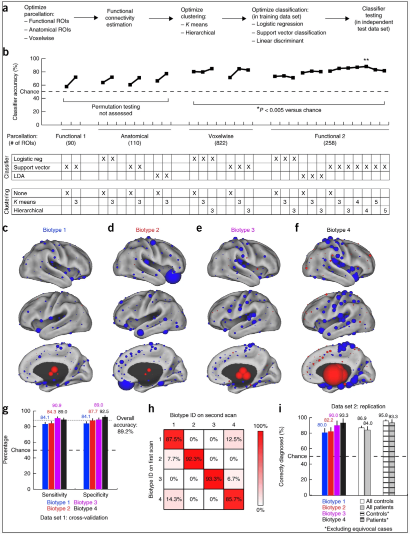
61k Accesses | **1348** Citations | **642** Altmetric | Metrics

☠☠☠ Clustering horror story ☠☠☠

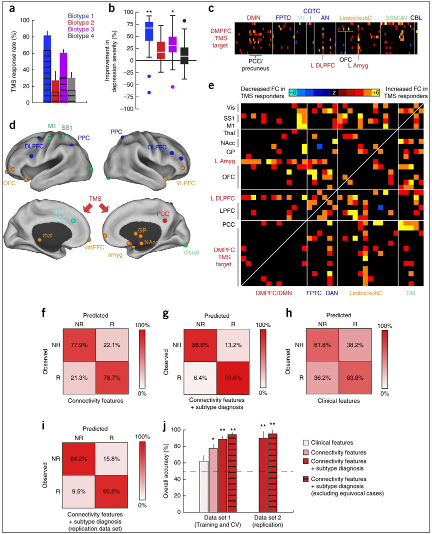



a
b
c
d
e
f
g
Biotype 1
Biotype 2
Biotype 3
Biotype 4
Abnormal connectivity (quartiles)
(first PC of shared abnormal connectivity features)
Prevalence:
Depressive symptom severity (z score)
Overall depression severity (total HAMD z score)
Anhedonia
Psychomotor retardation
Anxiety
Early insomnia
Middle insomnia
Anergia, fatigue

☠☠☠ Clustering horror story ☠☠☠



☠️☠️☠️ Clustering horror story ☠️☠️☠️



☠☠☠ Clustering horror story ☠☠☠

ELSEVIER

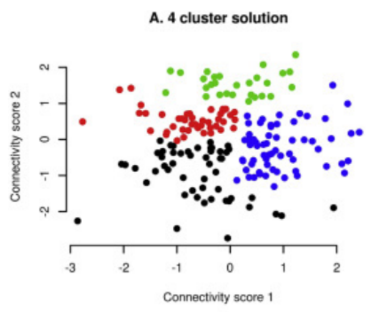
NeuroImage: Clinical

Volume 22, 2019, 101796

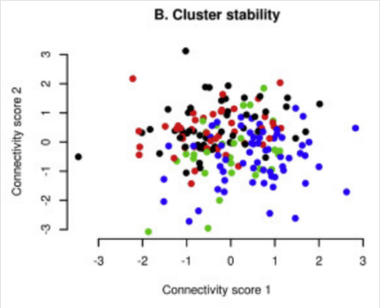
Evaluating the evidence for biotypes of depression: Methodological replication and extension of Drysdale et al. (2017)

Richard Dinga [a], Lianne Schmaal [b c], Brenda W.J.H. Penninx [a], Marie Jose van Tol [d], Dick J. Veltman [a], Laura van Velzen [a], Maarten Mennes [e], Nic J.A. van der Wee [f], Andre F. Marquand [e]

☠️☠️☠️ Clustering horror story ☠️☠️☠️



☠☠☠ Clustering horror story ☠☠☠



☠️☠️☠️ Clustering horror story ☠️☠️☠️

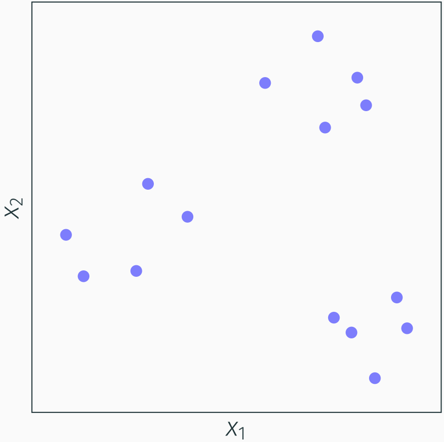
How can we avoid ending up with a clustering nightmare? 😱

- Quantitative evaluation of our clusters (see for instance Elements of Statistical Learning)
- Test cluster stability via cross-validation, half-split tests, bootstrap etc.
- Be cautious in interpretations

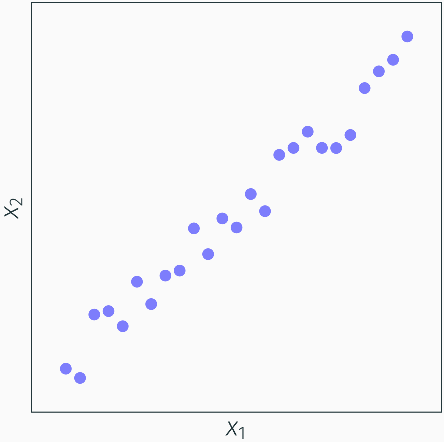
Dimensionality reduction

UNIVERSITY
OF OSLO

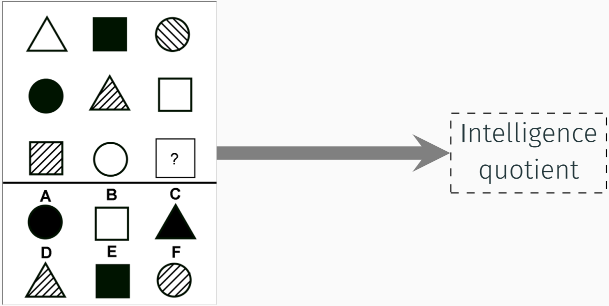
Dimensionality reduction: Motivation



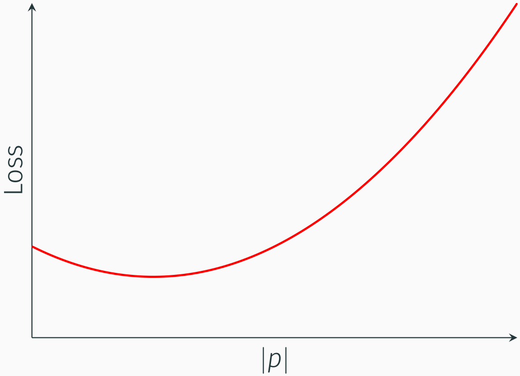
Dimensionality reduction: Motivation



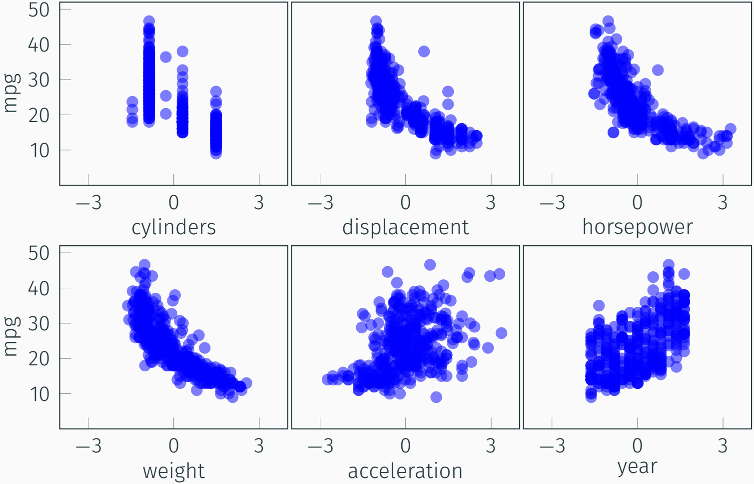
Dimensionality reduction: Motivation



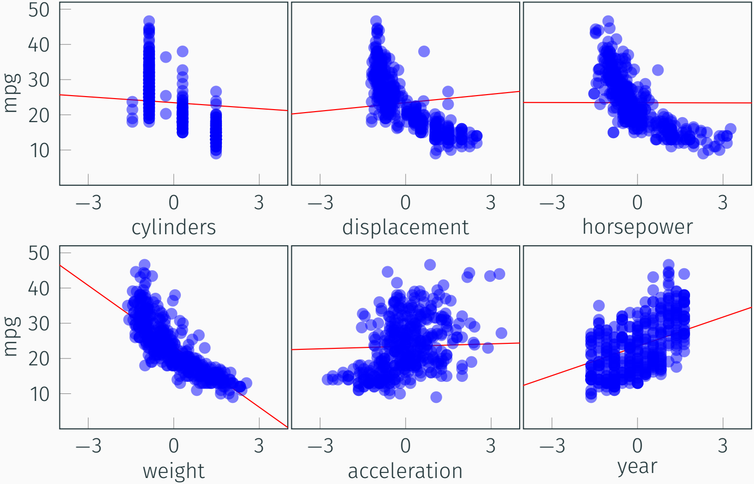
Dimensionality reduction: Motivation



Dimensionality reduction: Motivation



Dimensionality reduction: Motivation



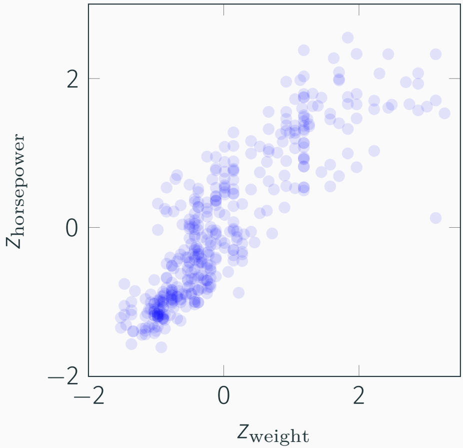
Dimensionality reduction: Motivation

1	0.30	0.86	0.89	0.41	0.93
0.30	1	0.41	0.34	0.29	0.36
0.86	0.41	1	0.84	0.68	0.89
0.89	0.34	0.84	1	0.50	0.95
0.41	0.29	0.68	0.50	1	0.54
0.93	0.36	0.89	0.95	0.54	1

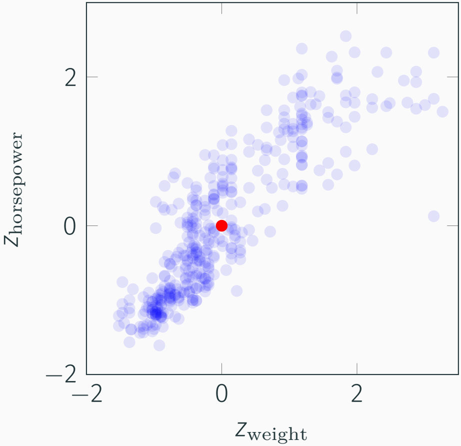
Dimensionality reduction: Motivation

1	0.30	0.86	0.89	0.41	0.93
0.30	1	0.41	0.34	0.29	0.36
0.86	0.41	1	$\boxed{0.84}$	0.68	0.89
0.89	0.34	0.84	1	0.50	0.95
0.41	0.29	0.68	0.50	1	0.54
0.93	0.36	0.89	0.95	0.54	1

Dimensionality reduction: Principal component analysis

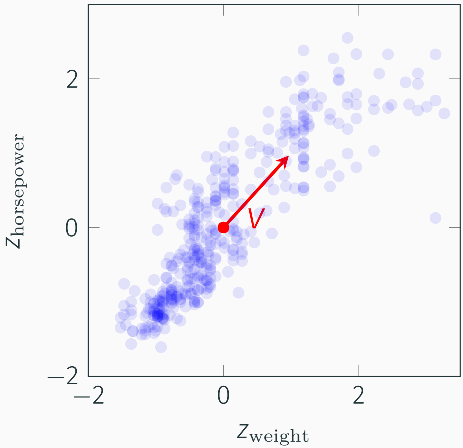


Dimensionality reduction: Principal component analysis



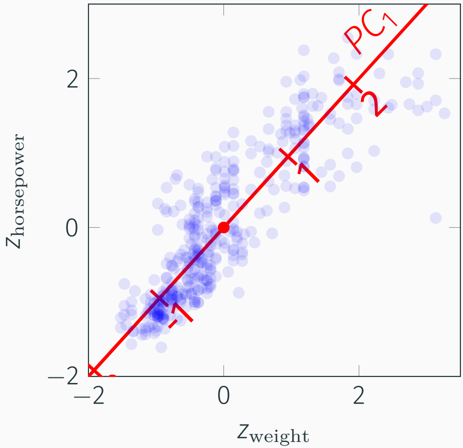
$c \rightarrow$ center of the data

Dimensionality reduction: Principal component analysis



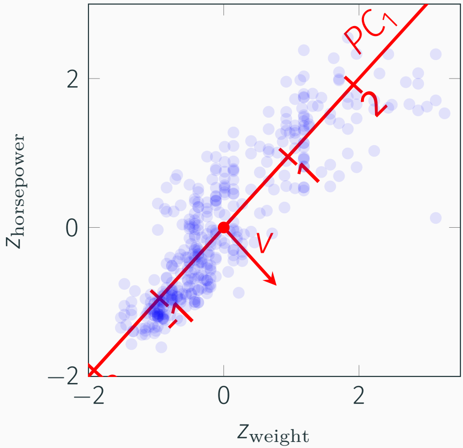
$v \rightarrow$ direction of maximum variance

Dimensionality reduction: Principal component analysis



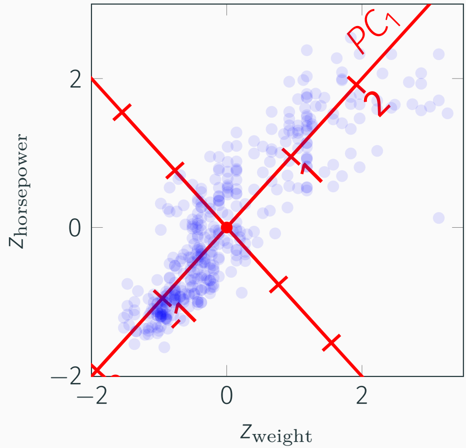
$$PC_1 \rightarrow 0.69 * z_{\mathrm{horsepower}} + 0.71 * z_{\mathrm{weight}}$$

Dimensionality reduction: Principal component analysis

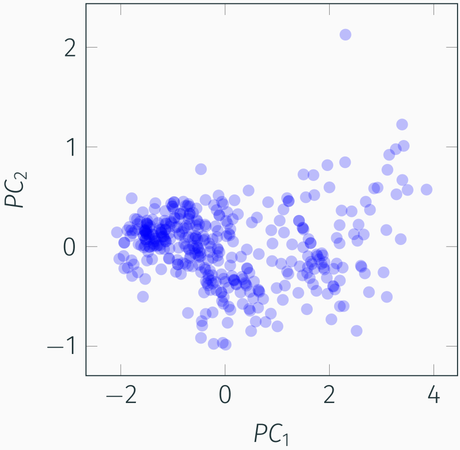


$v \rightarrow$ direction of maximum variance **orthogonal** to PC_1

Dimensionality reduction: Principal component analysis



Dimensionality reduction: Principal component analysis



Dimensionality reduction: Principal component analysis

mpg	horsepower	weight	PC1	PC2
18	130	3504	0.908	0.303
15	165	3693	1.709	0.517
18	150	3436	1.219	0.455
16	150	3433	1.217	0.457
17	140	3449	1.046	0.260
15	198	4341	2.856	0.583
14	220	4354	3.272	0.977

Dimensionality reduction: Principal component analysis

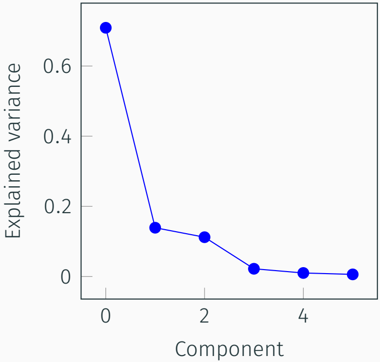
mpg	horsepower	weight	PC1	PC2
18	130	3504	0.908	0.303
15	165	3693	1.709	0.517
18	150	3436	1.219	0.455
16	150	3433	1.217	0.457
17	140	3449	1.046	0.260
15	198	4341	2.856	0.583
14	220	4354	3.272	0.977

$$PC_1 = 0.69 * z_{\text{horsepower}} + 0.71 * z_{\text{weight}}$$

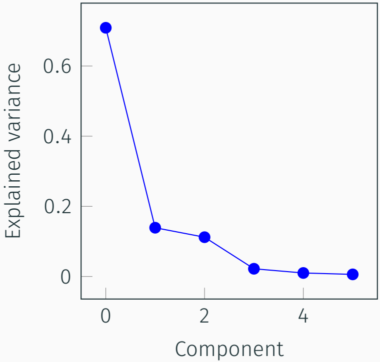
Dimensionality reduction: Principal component analysis

`http://localhost:8888/notebooks/notebooks%2FPCA.ipynb`

Dimensionality reduction: Principal component analysis

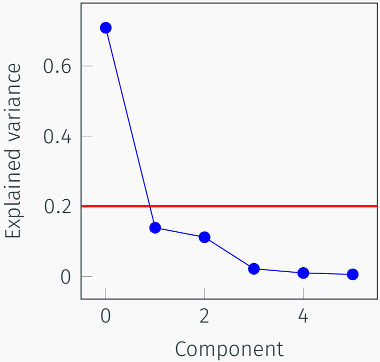


Dimensionality reduction: Principal component analysis



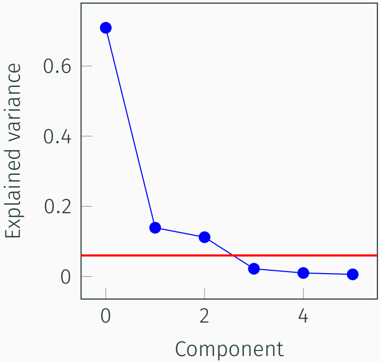
$$\hat{y} = \beta_0 + \sum_{i=0}^{n} \beta_i PC_i$$

Dimensionality reduction: Principal component analysis



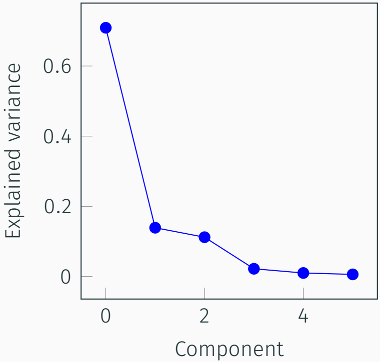
$$\hat{y} = \beta_0 + \sum_{i=1}^{1} \beta_i PC_i$$

Dimensionality reduction: Principal component analysis



$$\hat{y} = \beta_0 + \sum_{i=1}^{3} \beta_i PC_i$$

Dimensionality reduction: Principal component analysis



$$\hat{y} = \beta_0 + \sum_{i=0}^{n} \beta_i PC_i$$

n decided via a validation set, tested in a **held-out test set**

Dimensionality reduction: Principal component analysis

Principal component analysis: Transforms our dataset by computing *principal components* to replace our original variables.

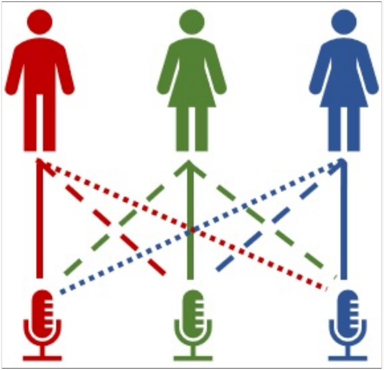
- Principal components are:
  - Linear combinations of the original variables
  - Orthogonal to each other, meaning that they capture different signals in our data (linearly uncorrelated)
- They can be useful for:
  - (Qualitatively) understanding the signal in our data
  - Reducing the number of predictors for modelling

Dimensionality reduction: Principal component analysis

Principal component analysis: Transforms our dataset by computing *principal components* to replace our original variables.

- Principal components are:
  - Linear combinations of the original variables
  - Orthogonal to each other, meaning that they capture different signals in our data (linearly uncorrelated)
- They can be useful for:
  - (Qualitatively) understanding the signal in our data
  - Reducing the number of predictors for modelling

Dimensionality reduction: Independent component analysis



Dimensionality reduction: Independent component analysis

Principal component analysis: Create orthogonal components that are linear combinations of our variables:

$PC_0 = \beta_0 x_0 + \beta_0 x_0 + \ldots + \beta_n x_n,$

$PC_1 = \gamma_0 x_0 + \gamma_1 x_1 + \ldots + \gamma_n x_n,$

$PC_0 \perp PC_1$ (e.g. no linear correlation)

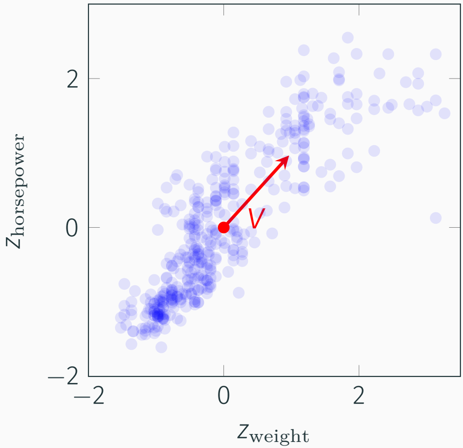
Independent component analysis: Represent each of our original variables as a linear combination of underlying sources:

$x_0 = \alpha_{00} s_0 + \alpha_{01} s_1 + \ldots + \alpha_{0n} s_n,$

$x_1 = \alpha_{10} s_0 + \alpha_{11} s_1 + \ldots + \alpha_{1n} s_n,$

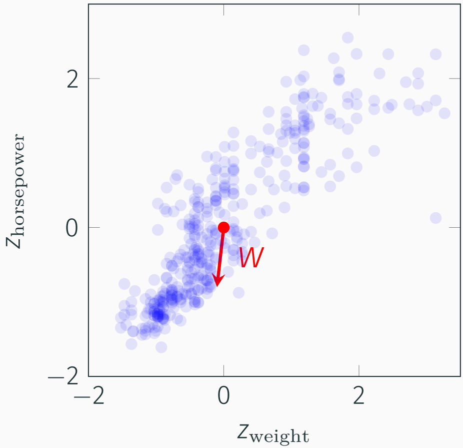
$s_0 \perp\!\!\!\perp s_1$ (e.g. statistically independent)

Dimensionality reduction: Independent component analysis



$v \rightarrow$ direction of greatest variance

Dimensionality reduction: Independent component analysis



$w \rightarrow$ direction that maximizes the non-Gaussanity of $w^T X$

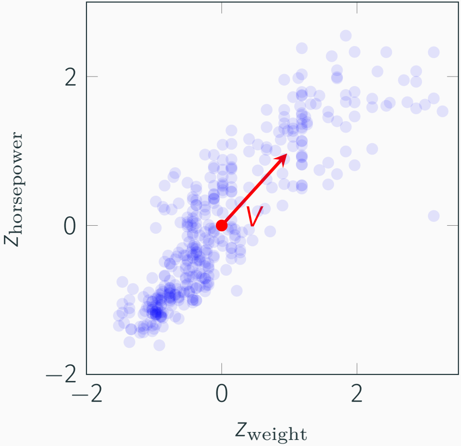
Dimensionality reduction: Independent component analysis

https://scikit-learn.org/dev/auto_examples/decomposition/plot_ica_blind_source_separation.html#sphx-glr-auto-examples-decomposition-plot-ica-blind-source-separation-py

Dimensionality reduction: Independent component analysis

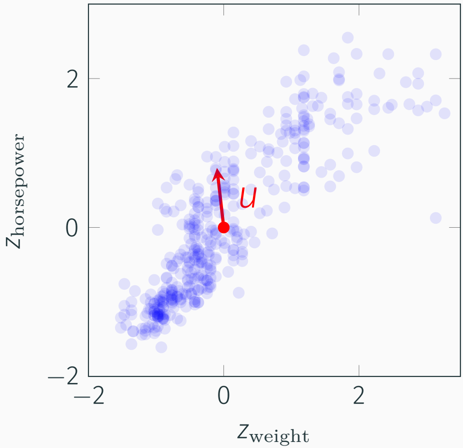
https://pmc.ncbi.nlm.nih.gov/articles/PMC7162660/

Dimensionality reduction: Partial least squares



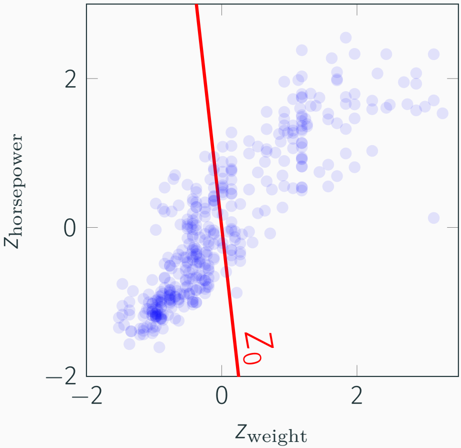
$v \rightarrow$ direction of maximum variance

Dimensionality reduction: Partial least squares



$u \to$ direction of greatest covariance between X and y

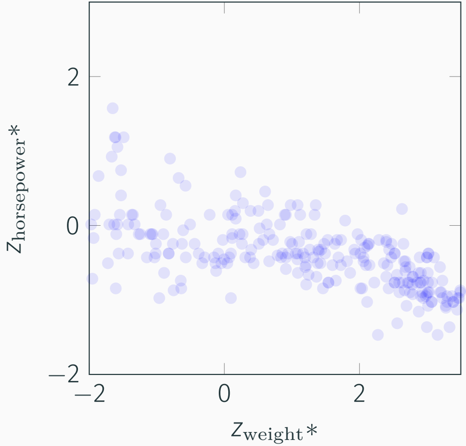
Dimensionality reduction: Partial least squares



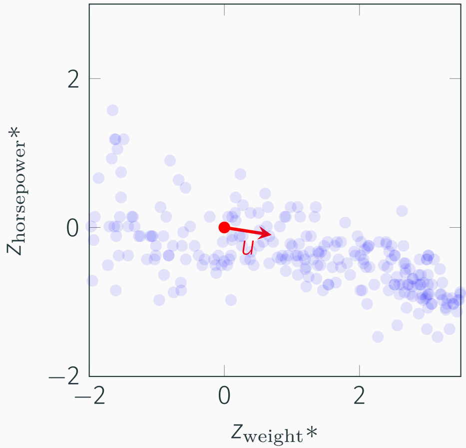
Dimensionality reduction: Partial least squares

$$z_{horsepower}* = z_{horsepower} - Z_0$$
$$z_{weight}* = z_{weight} - Z_0$$
$$y* = y - Z_0$$

Dimensionality reduction: Partial least squares



Dimensionality reduction: Partial least squares



$u \rightarrow$ direction of greatest covariance between $X*$ and $y*$

Dimensionality reduction: Partial least squares

Principal component analysis: Create orthogonal components that are linear combinations of our variables:
$PC_0 = \beta_0 x_0 + \beta_0 x_0 + \ldots + \beta_n x_n,$
$PC_1 = \gamma_0 x_0 + \gamma_1 x_1 + \ldots + \gamma_n x_n,$
$PC_0 \perp PC_1$ (e.g. no linear correlation),
that maximize the variance of X

Partial least squares: Create orthogonal components that are linear combinations of our variables:
$Z_0 = \beta_0 x_0 + \beta_0 x_0 + \ldots + \beta_n x_n,$
$Z_1 = \gamma_0 x_0 + \gamma_1 x_1 + \ldots + \gamma_n x_n,$
$Z_0 \perp PC_1$ (e.g. no linear correlation),
that maximize the covariance between X and y

Dimensionality reduction: Summary

Dimensionality reduction techniques allow us to reduce the number of variables in our dataset to either aid interpretation, or improve our models through implicit regularization.

- Principal component analysis (PCA): Finds components that are orthogonal and maximize variance
- Independent component analysis (ICA): Finds components that are non-Gaussian and statistically independent
- Partial least squares (PLS): Finds components that are orthogonal and maximize covariance between X and y