

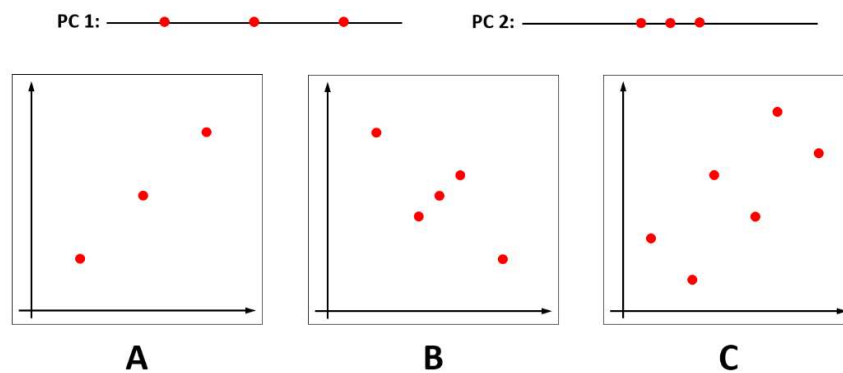
## Methods & Algorithms FS 2023

### Aufgabenblatt 3: Hauptkomponentenzerlegung

Die Bearbeitung der Aufgaben ist freiwillig; es erfolgt keine Bewertung.

#### Aufgabe 1

Gegeben sind folgende Datensätze A, B und C sowie die ersten beiden (nicht massstabsgetreuen) Komponenten PC 1 und PC 2 einer Hauptkomponentenzerlegung. Welche der drei Datensätze liefern mittels *Principal Component Analysis* beide Hauptkomponenten?



#### Aufgabe 2

Gegeben ist folgende prozentuale Verteilung der Varianz auf die Hauptkomponenten (PC) eines 8-dimensionalen Datensatzes gemäss *Principal Component Analysis*. Zeichnen Sie die Verteilung der Varianz in Python als kumulative Summe (d. h. aufsummiert) in ein Diagramm (x-Achse: Hauptkomponenten; y-Achse: Varianz [%]). Überlegen Sie ausserdem, auf wie viele Dimensionen sich der Datensatz praktisch ohne Verlust reduzieren lässt.

PC	1	2	3	4	5	6	7	8
Varianz (in [%])	80	18	1	0.25	0.25	0.25	0.25	0

#### Aufgabe 3

Der Iris-Datensatz aus dem Paket `sklearn.datasets` beinhaltet 150 Schwertlilien, die nach vier unterschiedlichen Merkmalen klassifiziert wurden. Laden Sie zunächst den Datensatz

`iris`, der u. a. die Teile `iris.data` (Grösse  $150 \times 4$ ) mit den Merkmalswerten sowie `iris.target` (Grösse  $150 \times 1$ ) mit den korrekten Klassifikationen enthält.

```
from sklearn import datasets  
iris = datasets.load_iris()  
x = iris.data  
y_true = iris.target
```

- a) Führen Sie mittels PCA eine Reduktion des vierdimensionalen Merkmalraumes auf zwei Dimensionen durch und visualisieren Sie das Ergebnis als Streudiagramm (Scatter Plot).
- b) Wie verteilt sich die Varianz des dimensionsreduzierten Datensatzes auf die beiden Hauptkomponenten? Ist das ein brauchbares Ergebnis?
- c) Bestimmen Sie mittels *k*-Mitten-Algorithmus die zugehörigen Cluster der dimensionsreduzierten Daten und visualisieren Sie das Ergebnis als Streudiagramm. Wie gut stimmt die vorhergesagte Klassifikation – rein optisch betrachtet – mit dem tatsächlichen Ergebnis aus a) überein?