DIMENSIONALITY REDUCTION

JENS BAETENS

WAT IS HET?

Dimensionality reduction technieken gaan proberen het aantal features dat nodig is te minimaliseren maar proberen om de behouden informatie te maximaliseren.

Veel gebruikte algoritmes:

- PCA of Principal Component Analyse
- Autoencoders
- Linear Discriminant Analysis
- Missing Values Ratio
- Low Variance Filter

WAAROM?

Curse of dimensionality -> Verscheidene technieken werken minder goed bij een hoog aantal features (performantie / trainingstijd)

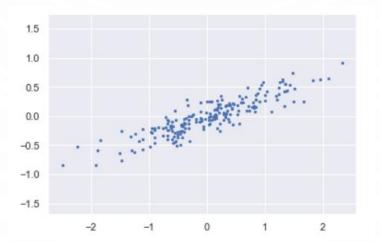
Data compressie

Ruis verwijderen

Visualisaties in 2D of 3D

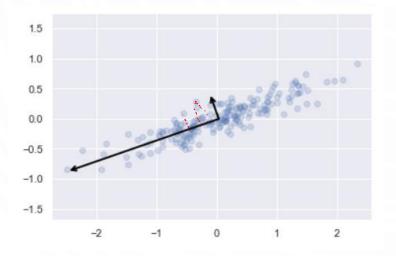
PCA – PRINCIPAL COMPONENT ANALYSE

Transformeren van assenstelsel voor de features om zoveel mogelijke informatie in de datapunten in de eerste features te krijgen zodat de andere weggelaten kunnen worden zonder belangrijke informatie te verliezen.



PCA – PRINCIPAL COMPONENT ANALYSE

2 features naar 2 principle components



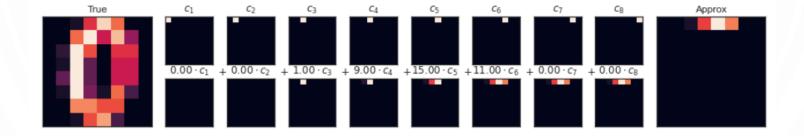


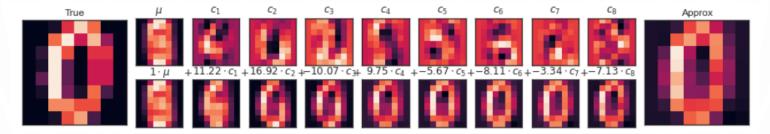
Voorbeeld van handgeschreven cijfers (scaled MNIST)

Afbeeldingen van 8x8 pixels met cijfers van 0 t.e.m. 9 (64 features)



Doel van PCA: Reduceer 64 features maar behoud zoveel mogelijk van de informatie

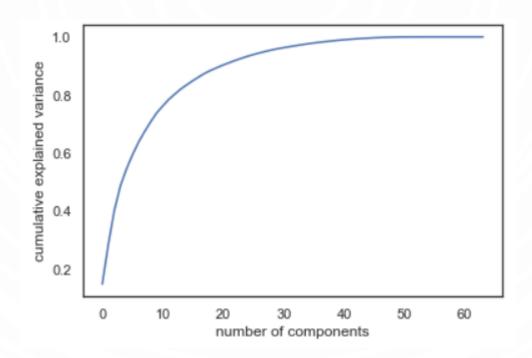




Na Pca geven de eerste 8 features al een goede benadering van het origineel

- Sterke compressie van de data (factor 8)
- Verbetert performantie van de modellen

In computer visie worden deze principal components ook Eigenfaces genoemd

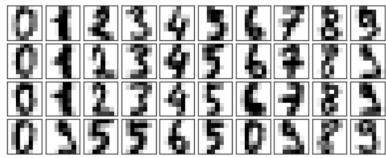


PCA VOOR NOISE FILTERING

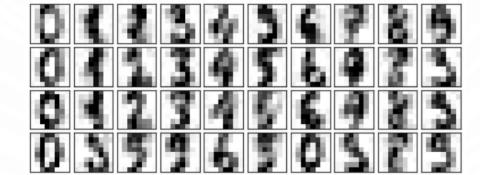
Aangezien de ruis willekeurig is gaat PCA vooral focussen op de zaken die gemeenschappelijk is tussen de beelden van dezelfde klasse.

Hierdoor wordt de ruis eruit gefilterd

PCA VOOR NOISE FILTERING







Na PCA