

Visualisierung von Musikdaten mittels t-SNE und PCA am Beispiel pgvector

Jannis Gehring
INF22B, Data Warehouse
Duale Hochschule Baden-Württemberg (DHBW)
Stuttgart, Deutschland
inf22115@lehre.dhbw-stuttgart.de

Abstract—

CONTENTS

I) Grundlagen	1
I.A) Principal Component Analysis (PCA)	1
I.B) t-SNE	1
I.C) comparison	1
II) Installation	1
III) Umsetzungsbeispiel	1
References	1

I. GRUNDLAGEN

A. Principal Component Analysis (PCA)

Principal Component Analysis (PCA) ist ein lineares Verfahren zur Dimensionsreduktion. Es wurde in der ersten Hälfte des zwanzigsten Jahrhunderts entwickelt, fand aber aufgrund seiner Berechnungsanforderungen erst in den 60ern breite Anwendung. [1]

Mathematisch liegt Principal Component Analysis (PCA) eine Eigenvektorberechnung zugrunde. Vereinfacht wird zu Beginn die Kovarianz-Matrix S des Datensatzes berechnet. Dann werden die Eigenvektoren dieser Matrix berechnet und nach dem Betrag ihrer jeweiligen Eigenwerte sortiert. Von diesen n Eigenvektoren werden nun, je nach Anwendungsfall, die k ersten gewählt ($k \leq n$). Mit einer weiteren Matrix W , die jene gewählten Eigenvektoren als Zeilenvektoren hat, wird nun die Transformation der Ausgangstupel in den (meist niedriger-dimensionalen) Raum durchgeführt. [2]

Das folgende Beispiel gibt dieser mathematischen Definition eine handfeste Intuition. Hier entspricht PCA dem iterativen Auswählen von zueinander orthogonalen Linien durch den Datensatz, die diesen am besten teilen. *Am besten* bedeutet hier, dass die Varianz der Projektionen auf diese Linie maximal sind. [2]

PCA illustratives Beispiel

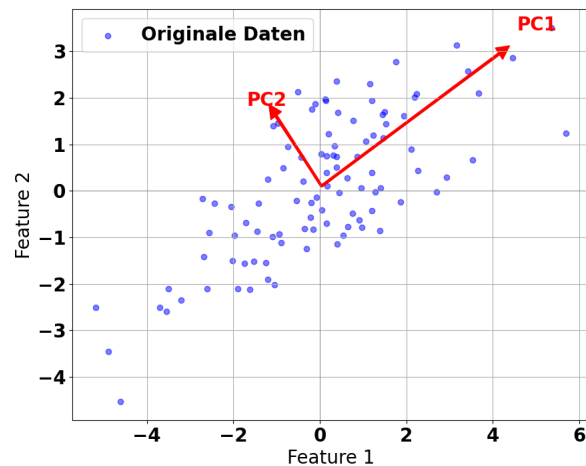


Fig. 1: Veranschaulichung von PCA mit $n = k = 2$

B. t-SNE

C. comparison

II. INSTALLATION

III. UMSETZUNGSBEISPIEL

REFERENCES

- [1] A. Maćkiewicz and W. Ratajczak, "Principal components analysis (PCA)," *Computers & Geosciences*, vol. 19, no. 3, pp. 303–342, 1993, doi: [https://doi.org/10.1016/0098-3004\(93\)90090-R](https://doi.org/10.1016/0098-3004(93)90090-R).
- [2] M. E. Tipping and C. M. Bishop, "Mixtures of probabilistic principal component analysers," *Neural Computation* 11(2), pp. 443–482, 1999.