

Hauptkomponentenanalyse

-

Principal Component Analysis (PCA)

1 Aufgaben

1.1 PCA mit Digits Datensatz

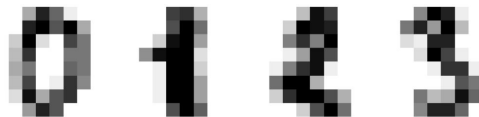


Abbildung 1: Digits Datensatz

- (a) Erstellen Sie ein neues Python-Skript.
- (b) Laden Sie den Digits-Datensatz von scikit-learn.
- (c) Importieren Sie die PCA-Klasse und führen Sie eine Hauptkomponentenanalyse mit $n_components = 2$ mit den Trainingsdaten durch.
- (d) Erstellen Sie einen Scatter-Plot mit den neuen Features.
- (e) Ändern Sie die PCA nun so, dass die ersten drei Hauptkomponenten analysiert werden und plotten Sie diese in einem 3D-Plot.
- (f) Führen Sie eine komplette PCA durch und plotten Sie die prozentuale erklärte Varianz der Hauptkomponenten in einem neuen Fenster. Nutzen Sie dafür das Attribut *explained_variance_ratio_*.

1.2 PCA mit CIFAR10-Datensatz

Nutzen Sie für diese Aufgabe die Datei *cifar10_bearbeitet.py* aus dem Modul Machine Learning I.

- (a) Importieren Sie die PCA-Klasse und führen Sie eine Hauptkomponentenanalyse mit $n_components = 32$ auf den Trainingsdaten durch.
- (b) Transformieren Sie die Trainings- und Testdaten auf die neuen Features.
- (c) Führen Sie die Klassifizierung mit den neuen Eingabedaten durch. Vergleichen Sie die Genauigkeit mit und ohne PCA.
- (d) Messen Sie die Zeit, die zur Ausführung des Skripts benötigt wird und vergleichen Sie diese mit und ohne PCA.