

Hauptkomponentenanalyse

-

Principal Component Analysis (PCA)

1 Aufgaben

1.1 PCA mit Digits Datensatz

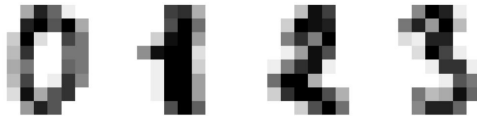


Abbildung 1: Digits Datensatz

- Erstellen Sie ein neues Python-Skript.
- Laden Sie den Digits-Datensatz von scikit-learn.
- Importieren Sie die PCA-Klasse und führen Sie eine Hauptkomponentenanalyse mit $n_components = 2$ mit den Trainingsdaten durch.
- Erstellen Sie einen Scatter-Plot mit den neuen Features.
- Ändern Sie die PCA nun so, dass die ersten drei Hauptkomponenten analysiert werden und plotten Sie diese in einem 3D-Plot.
- 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 und k-Nearest-Neighbours

Für diese Aufgabe müssen Sie den k-Nearest-Neighbour-Klassifikator aus dem Modul Machine Learning I verwenden.

- (a) Trainieren Sie zunächst *ohne PCA* einen kNN-Klassifikator auf den *ersten 50* Bildern des Digits-Datensatzes.
- (b) Wenden Sie den trainierten Klassifikator nun auf die *übrigen 1747 Bilder* an und berechnen Sie die erzielte Genauigkeit.
- (c) Wenden Sie als Zwischenschritt nun eine PCA an, bevor Sie den kNN-Klassifikator verwenden.
- (d) Wie verändert sich die Genauigkeit des Klassifikators, wenn Sie die Anzahl der PCA-Komponenten (`n_components`) variieren?