

Machine Learning (Online)

Übung 3

Ziel dieser Übung ist das Clustering von Immobilien mit Interpretation der entstandenen Cluster und die Dimensions-Reduktion von Features. Wir verwenden dazu wiederum den vollen Datensatz `house_data.csv` der Übung 1.

Aufgaben:

- 1) **(10 Punkte)** Clustern Sie die Immobilien indem Sie geeignete Features dafür suchen und allenfalls sinnvoll normalisieren. Erstellen Sie einen *Elbow*-Plot für $2 \leq k \leq 20$ und wählen Sie eine geeignete Cluster-Anzahl k . Führen Sie ein Clustering durch für die gewählte Cluster-Anzahl k , visualisieren Sie die entstandenen Cluster mit geeigneten Darstellungen und versuchen Sie, die Cluster zu charakterisieren (welche Eigenschaften haben die Gebäude im jeweiligen Cluster ?)

Tipp 1: Eine Möglichkeit, einen Cluster zu charakterisieren, besteht darin, dass Sie untersuchen, welche Features die kleinste Varianz haben (am ähnlichsten sind).

Tipp 2: Neben der *Elbow*-Methode gibt es noch andere Varianten („die Sie kennen sollten“) zum Bestimmen einer geeigneten Cluster-Anzahl :

<http://www.sthda.com/english/wiki/determining-the-optimal-number-of-clusters-3-must-known-methods-unsupervised-machine-learning>

- 2) **(15 Punkte)** Unterziehen Sie das in Aufgabe 1 gewählte Modell einer *Principal Component Analysis* und clustern Sie die Immobilien diesmal auf den ersten 3 *Principal Components*. Erstellen Sie wiederum einen *Elbow*-Plot für $2 \leq k \leq 20$ und bestimmen Sie die geeignetste Cluster-Anzahl. Führen Sie mit der gefundenen Cluster-Anzahl ein Clustering durch und charakterisieren Sie die entstandenen Cluster.

- Diskutieren Sie den Unterschied der Cluster von Aufgabe 1 und Aufgabe 2.
- Visualisieren Sie die ersten 3 *Principal Components* und suchen Sie nach Eigenschaften, welche das Clustering negativ beeinflussen könnten.
- Genügt es, nur die ersten 3 *Principal Components* zu verwenden ? Begründen Sie Ihre Antwort.

Abgabe : Beschreiben Sie Ihre Lösung/Modelle in einem PDF und senden Sie dieses zusammen mit dem Code als Zip-Datei in der Form

Name_Vorname.zip

bis spätestens 4. Juni 2018 an lukas.neukom@fhnw.ch.