

機器學習導論

Introduction to Machine Learning

Homework 4 Dimensionality Reduction & Clustering

(Thursday, May 14th, 2021)

”InterestsSurvey.csv”這筆資料是一份來自於不同族群的興趣分析問卷調查，其中的[‘group’]欄位表示不同的族群，但並不代表這些問卷的實際分類，請結合課堂所學對這筆問卷調查結果進行社群聚類分析。

1. Data pre- processing (10%):

請以圖表呈現 data 中你認為必須進行前處理的部分，以文字詳述前處理的方法與過程，最後再以圖表呈現前處理的結果。

2. Dimensionality reduction (25%):

請以文字詳述說明你選擇降維方法的依據及理由，包含參數配置，配以表格或圖佐證。

3. Clustering (25%):

請以文字詳述說明你選擇聚類方法的依據及理由，包含參數配置，配以表格或圖佐證。

4. Result Interpretation (30%):

請解釋你的社群聚類分析結果，你認為這些結果有什麼樣的涵義，請盡可能的詳細說明，以圖或表格方式呈現。

5. Tidy up your report (10%):

請確保你的報告排版整齊、文句通順、用詞精準、用字正確，中文字型請用標楷體，英文字型請用 Time News Roman，內文字型大小 12 標題字型大小 14，行距為單行間距。

Hint:

本次作業並無標準解答，請各位同學盡可能地賦予你的分析結果意義，在做每一個步驟時，都請先捫心自問自己做這件事的原因為何，結果為何，用客觀的數據去佐證你主觀的判斷，Good luck!