

INFS4203/7203 Assignment 3

Semester 2, 2020

Marks:	6 marks
Due Data:	23:59, 11 Oct 2020, Brisbane time
What to submit:	a report in the pdf format
Where to submit:	Turnitin link via blackboard

The goal of this project is to gain practical experience in applying clustering to real data. You must work on this project **individually**. The standard academic honesty rules apply. You are required to read and understand the School Statement on Misconduct, available on the School's website at:

<http://ppl.app.uq.edu.au/content/3.60.04-student-integrity-and-misconduct>

Dataset:

We have a dataset stored in **data.csv**, which contains 1797 data points (1797 rows) and 2 attributes (2 columns). The first attribute is named "x". The second attribute is named "y".

Task:

Given the dataset, choose a proper algorithm to do clustering. Justify the result from the following aspects (maximum 500 words):

1. Why this algorithm is suitable for this dataset? [3 marks]
2. How do you determine the number of clusters or other parameters/Python function arguments? [3 marks]

Visualization (e.g., Fig. 1) is suggested to help explain your arguments.

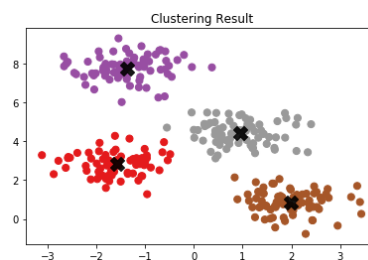


Fig. 1