資料分析方法-HW8

經研一 江彥亨 R11323040

Q1.

Hierarchical clustering :

	Woman_pred	Man_pred
Woman	19	21
Man	176	184

Kmeans :

	Woman_pred	Man_pred
Woman	30	10
Man	188	172

Based on the results, neither Hierarchical clustering nor Kmeans perfectly match the two genders and their respective accuracy are almost the same. In addition, Kmeans has better prediction for men, but worse prediction for women.

Q2.

Hierarchical clustering :

	1_pred	2_pred	3_pred
1	129	88	28
2	3	27	38
3	0	26	53

Kmeans :

	1_pred	2_pred	3_pred
1	89	60	96
2	1	50	17
3	0	70	9

DBSCAN :

	outliers	1_pred	2_pred	3_pred
	0	0	0	0
1	2	228	10	5
2	0	68	0	0
3	2	77	0	0

Based on the results, all three algorithms don’t perfectly match the origin. Although DBSCAN has the highest accuracy, it can only classify samples with 'origin=1' and fails to classify the other two classes. On the other hand, Kmeans performs well in predicting samples with 'origin=2', but its performance is poor for the other two classes. However, the performance of Hierarchical clustering is relatively good compared to the other two methods. Therefore, Hierarchical clustering is a relatively good method for this dataset and produces similar results to the supervised learning methods.