Michael Han 101157504

Azamat (Rush) Galimov 101263850

COMP3105A

Assignment 3

Table 1: Training accuracies with different number of training dataset sizes

n	Model 1	Model 2
16	0.8625	1
32	0.921875	0.990625
64	0.8765625	0.946875
128	0.86640625	0.92734375

```
1d) Training Accuracy Matrix (4 x 2):
[[0.8625     1.        ]
 [0.921875   0.990625  ]
 [0.8765625  0.946875  ]
 [0.86640625 0.92734375]]
```

Table 2: Test accuracies with different number of training dataset sizes

n	Model 1	Model 2
16	0.7076	0.8862
32	0.8098	0.8703
64	0.8187	0.9012
128	0.8393	0.9185

```
1d) Testing Accuracy Matrix (4 x 2):
[[0.7076 0.8862]
 [0.8098 0.8703]
 [0.8187 0.9012]
 [0.8393 0.9185]]
```

Table 3: Training accuracies with different number dimensions

Dim k	Model 1	Model 2
1	0.84851562	0.85945313
2	0.65054687	0.94257813

```
2d) average train acc
[[0.84851562 0.65054687]
 [0.85945313 0.94257813]]
```

Table 4: Test accuracies with different number dimensions

Dim k	Model 1	Model 2
1	0.84216	0.63107
2	0.83815	0.91697

```
2d) average test acc
[[0.84216 0.63107]
 [0.83815 0.91697]]
```

Table 5: Objective values for different k

k	2	3	4	5	6	7	8	9
obj _val	2.35507252 9689218	1.39570518 31072367	0.74704928 03163796	0.56559062 47538163	0.51484050 34384456	0.46287470 28814755	0.43250155 12513061	0.344554488 20676876

```
obj_val:  [2.355072529689218, 1.3957051831072367, 0.7470492803163796, 0.5655906247538163, 0.5148405034384456, 0.4628747028814755, 0.4325015512513061, 0.34455448820676876]
```

I think $k = 4$ is appropriate for this dataset, because the data is generated with 4 distinct clusters. $k = 2$ and $k = 3$ have high objective values, and the improvement for the objective value when $k \geq 5$ is very small.

2b) results