ML2 DSEB W4 - GMM

Nguyen Tuan Duy - 11204971

February 2023

1 Problem 3

Pros and cons of GMM:

	GMM	K-means
Pros of GMM	Useful for data points that fall at the border of two clusters (GMM is a soft clustering method)	Not possible (K-means is a hard clusterin
	Less sensitive to scale	More sensitive to data of var
	Don't assume spherical clusters	assume spherical clusters
Cons of GMM	Cannot work well with very irregular shape	Same
	Comparatively slow	Faster (performance scales linearly
	Manually specify number of clusters	Same
	Sensitive to outliers	Same (more sensitive)
	Sensitive to initialization conditions (Starting point)	Same
	More complicated to implement and understand	Simpler

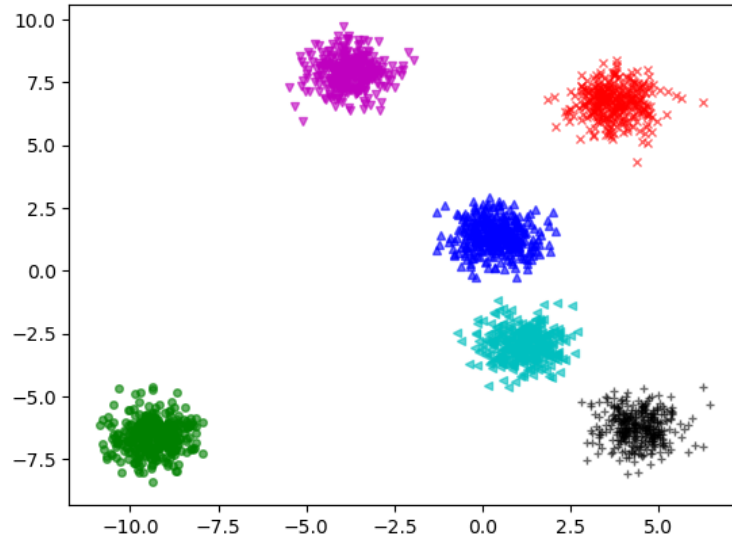# 2 Problem 4

Example 1: data of varying sizes and density.

Figure 1: Original data

The original data contains 6 blobs varying in distances between blobs(green blob is very far from other blobs whereas blue blob is very close to cyan blob)
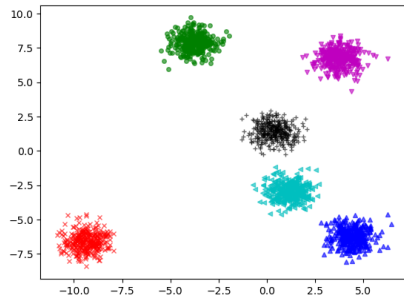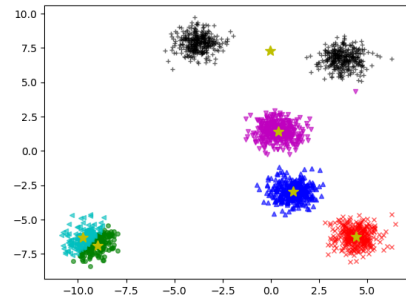

Figure 2: result from GMM

Figure 3: result from K-means

K-means has trouble clustering data where clusters are of varying sizes whereas gmm have no trouble.

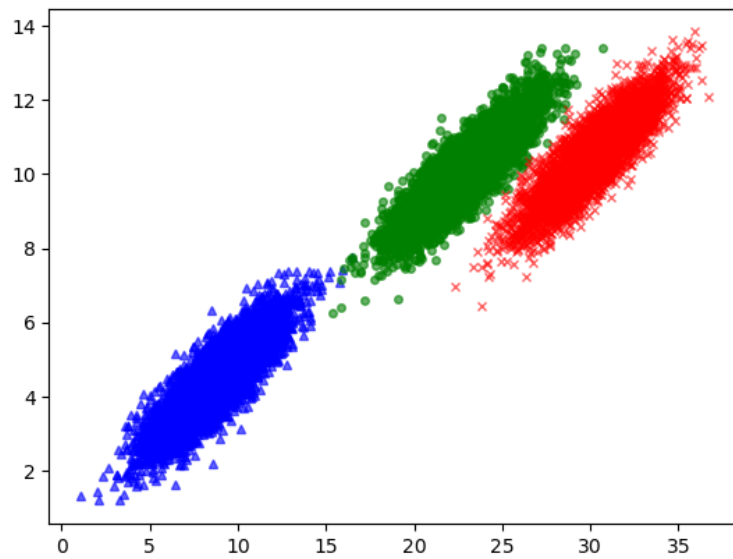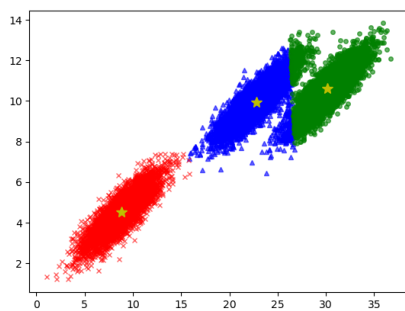Example 2: data with high covariance
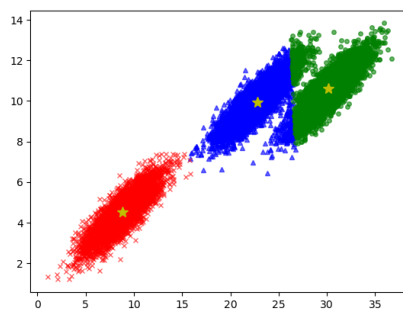
Figure 4: Original data

Figure 5: result from GMM

Figure 6: result from K-means

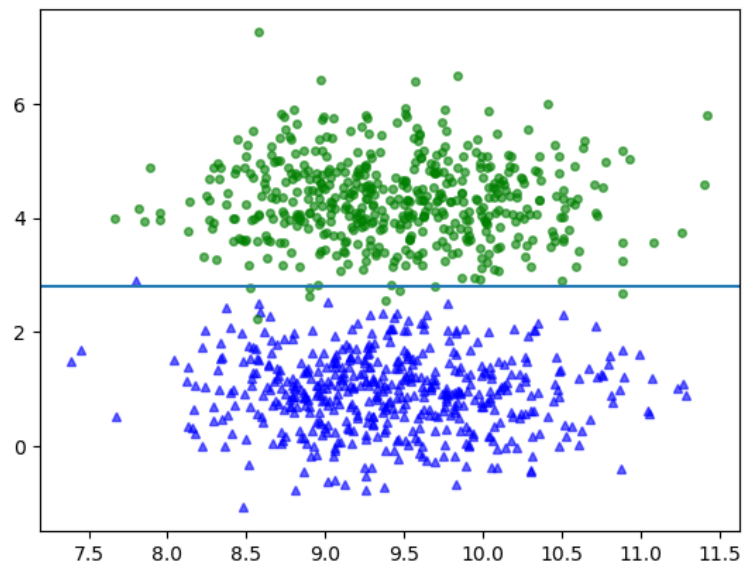Expample 3: datapoints fall at the border of two clusters:

Figure 7: some points are near decision boundary

In this case it is better to use soft clustering method (like GMM) rather than (K-means)