# Report

## Pros/Cons of Clustering Methods

K Means Clustering

Pros

* Easy to implement.
* Guaranteed convergence.
* Scales to large datasets

Cons

* a negligent edge of each cluster, because the priorities are set on the center of the cluster, not on its borders. (Because optimizing k-means equivalent to optimizing MLE for mixed Gaussian model with covariance matrix )
* an inability to create a structure of a dataset with objects that can be classified to multiple clusters in equal measure.
* a need to guess the optimal *k*number, or a need to make preliminary calculations to specify this gauge.
* NP-hard.
* Handling of outliers.
* Requires initialization of centers.
* Difficulty clustering data of varying densities and sizes
* Curse of dimensionality- with increasing amount of features the distances between points obtain a lower ratio of std. variation to mean.

Mixed Gaussians Model

Pros

* Unlike the centroid-based models, the EM algorithm allows the points to classify for two or more clusters – it simply presents you the possibility of each event, using which you can conduct further analysis.
* The borders of each cluster compose ellipsoids of different measures unlike k-means, where the cluster is visually represented as a circle.

Cons

* The algorithm simply would not work for datasets where objects do not follow the Gaussian distribution. It is more applicable to theoretical problems rather than the actual measurements or observations.
* Insufficient data leads to overfitting and difficulty estimating the covariance matrices which leads to divergence

Spectral Clustering

Pros

* Can recognize clusters that do not have a clear blob-shape, such as concentric rings.
* Polynomial time optimizable for fixed kfor cost function G\_cut (defined in lecture)
* Avoids curse of dimesionality- projects data onto lower dimensional space and then perform clustering using k-means (or other method)

Cons

* tends to produce small clusters
* NP hard for cost function G\_cost\_cut for k=2 (defined in lecture)