***Definitions and Explanations:***

**K-means:** Minimizes the distance between each point with the mean of each points cluster assigned. *Pros:* Fast, easy to implement. *Cons*: Local optimal, limited to spherical distance.

**Graph Clustering**: ratio-cut; minimizes the weight of edges cut over the size of each side (clusters) of that cut, normalized-ratio-cut; minimizes the weight of edges cut over all the edge weights in a cluster including the cut weights. *Pros:* Good for when clusters aren’t the same size. *Cons:* slow.

**EM Clustering**: maximizes likelihood of each point in D to a set of clusters (Gaussians) defined by theta the paremeter {meani, covariancei P(Ci)}. *Pros:* Very accurate, includes probability instead of discrete values. *Cons*: Slow, bad for concave data.

**LDA**: Finds the direction which maximizes the projected means while minimizing scatter.

**Two Main Concepts of SVM:** 1. Maximizes margin by creating a hyperplace h(x)=0 s.t. we maximize the margin between all support vectors and the hyperplane. 2. Use of Kernels; allows for a non-linear or better similarity function to be leveraged with an understanding of the data set.

**Parts of SVM That Line Up With Its Goals:** the max w ½ wTw piece maximizes the margin, while the max zeta piece allows for the most extra points inside the margin, while max b pieces tells us we want the above optimization to have the best placed plane to do so in classifying new points.

**Dual Formulation Importance:** Use of the Kernel function.

**DBSCAN**: Good for images/ removes noise points. Single Link in KNN will be good for images of convex nature. **Hypothesis Set**: The larger it is the more variance error. **Training Error** = bias error + variance error + noise.

***Preventing Over-Fitting*: KNN:** Over-fits at low K. Increase K to manage overfitting. **Decision Trees:** Over-fits when the classifications are perfect between each axis parallel cut, can be managed by allowing an error or mixture of classes on either side of the cut. **SVM**: Over-fits by making a super narrow padding 1/||w||, this can be managed by increasing the *c* hyper-parameter to allow more bias error of allowing points inside of the margin (allows intentional misclassification). So increase C.