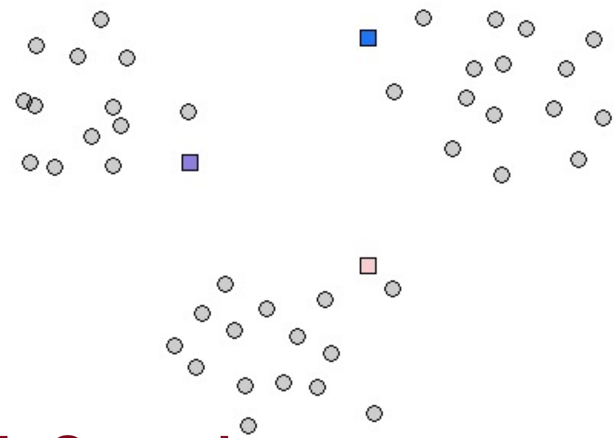
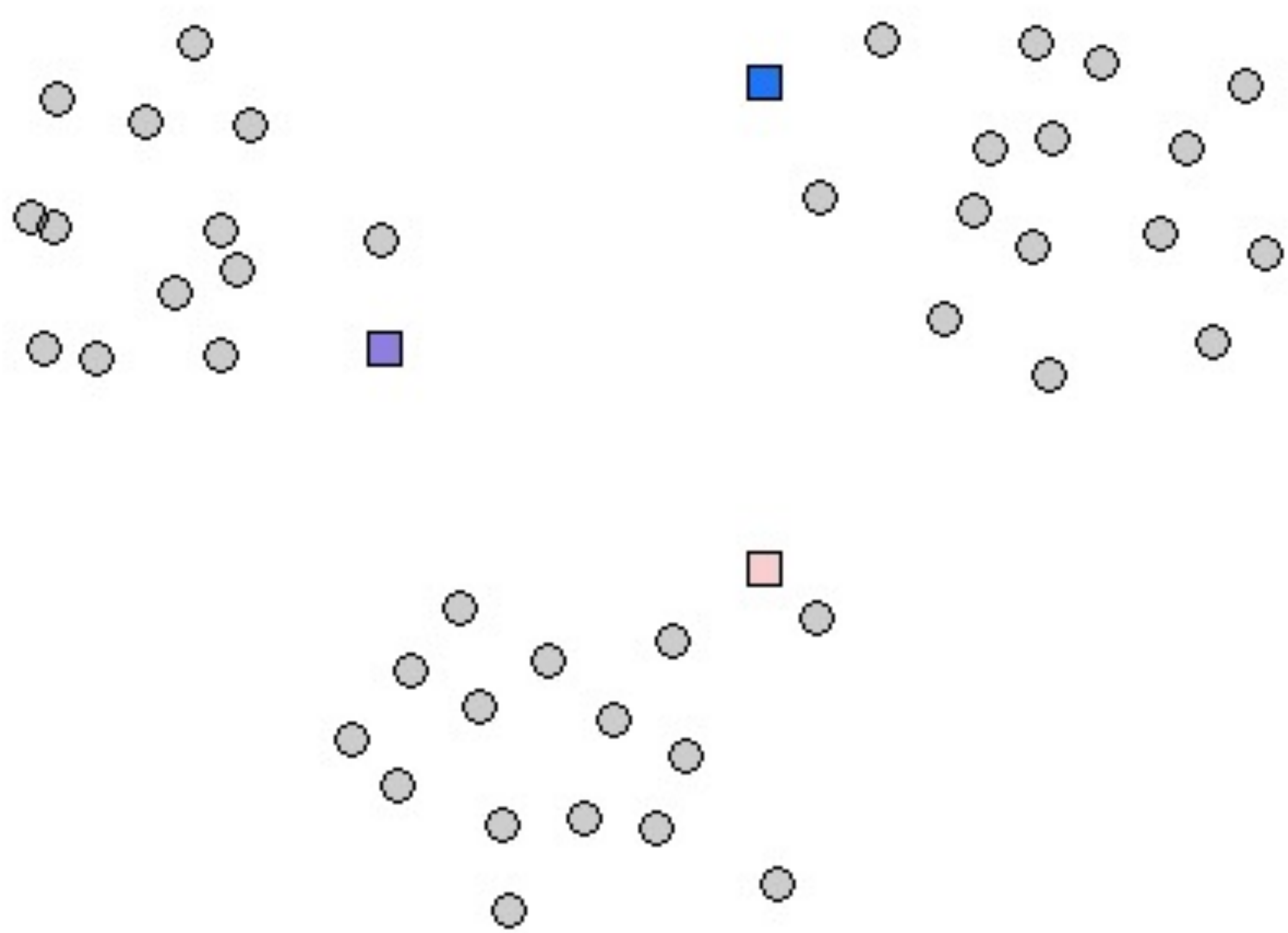
K-means

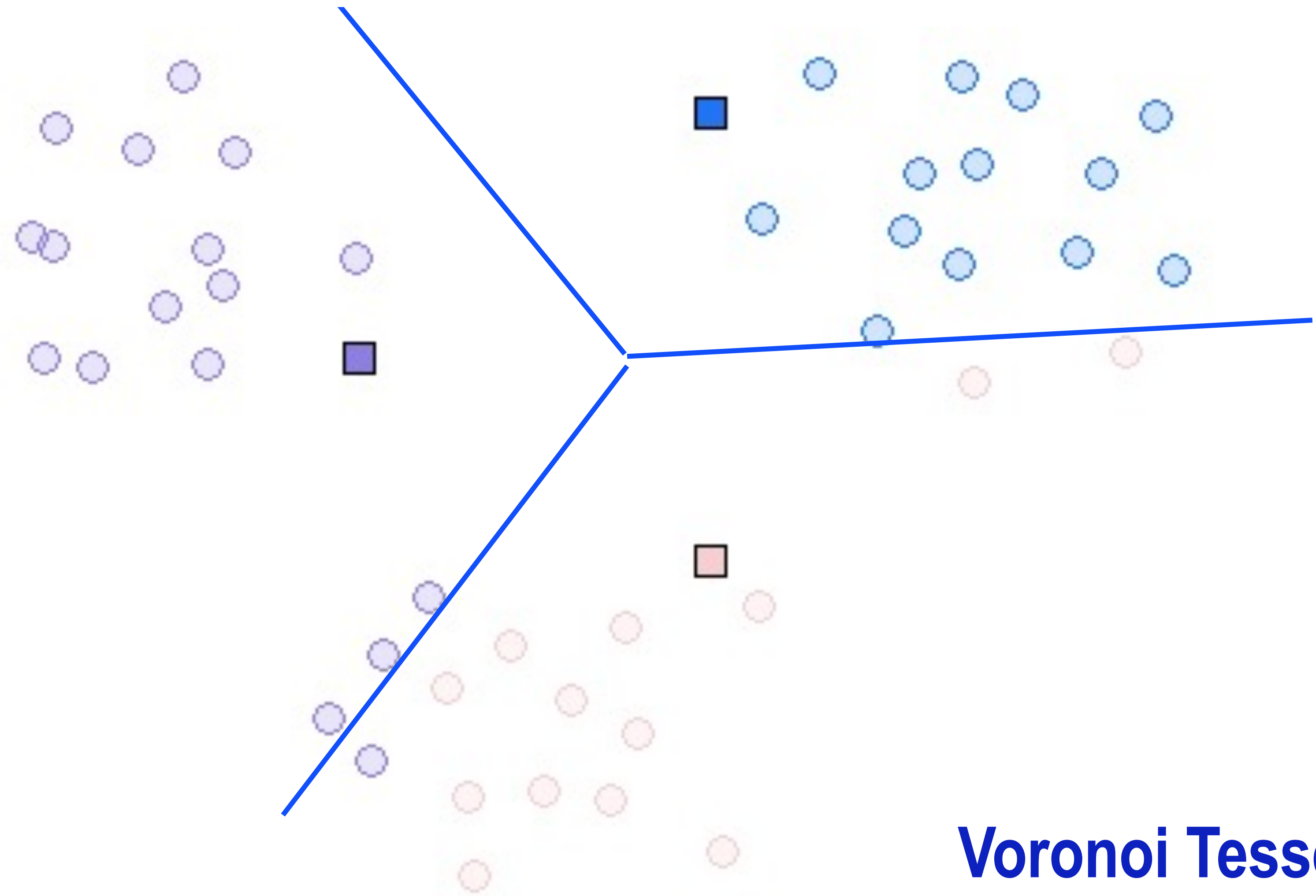
Lyle Ungar



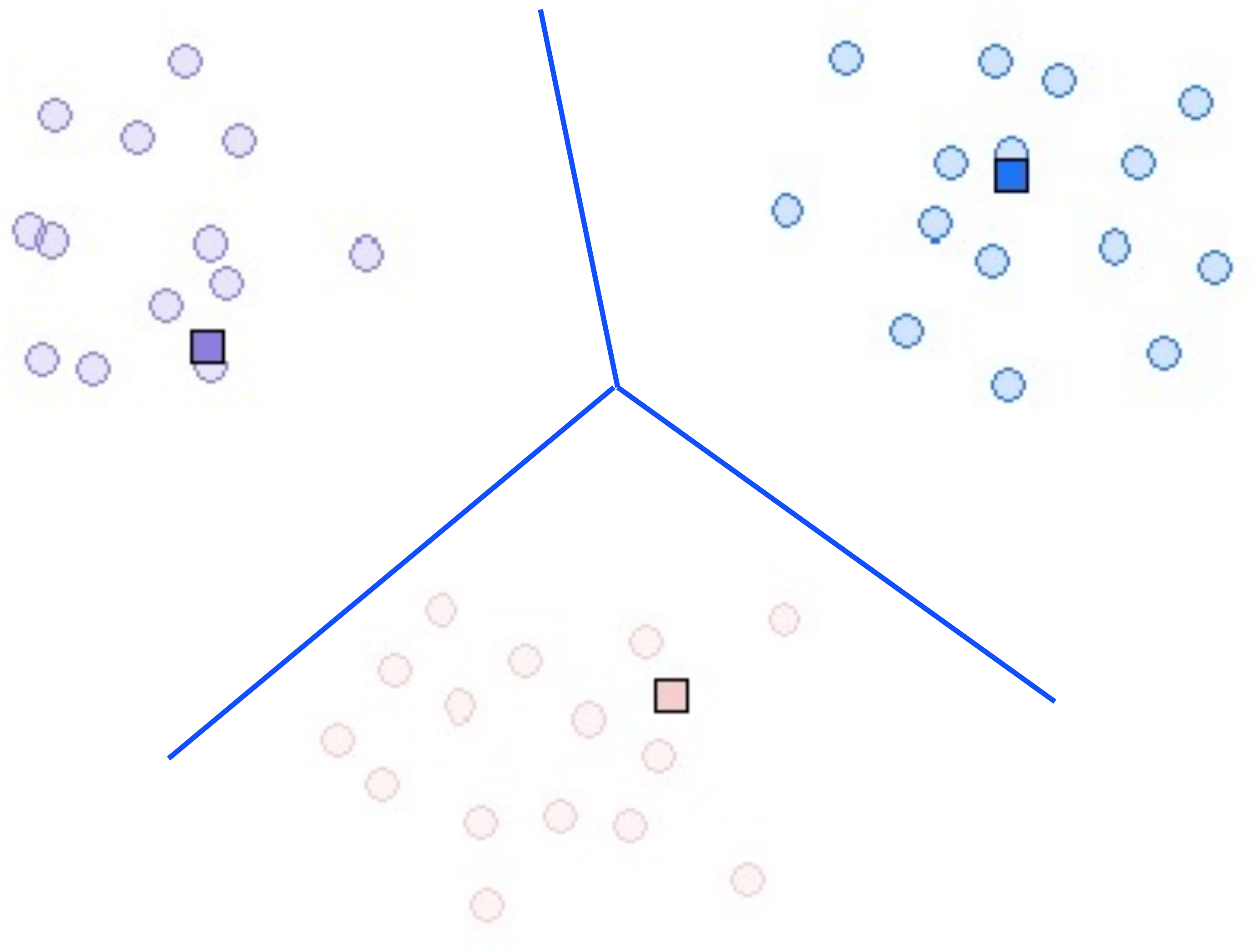
Why cluster?
K-means: loss function, algorithm
Relation to EM and Gaussian Mixture Models (GMM)

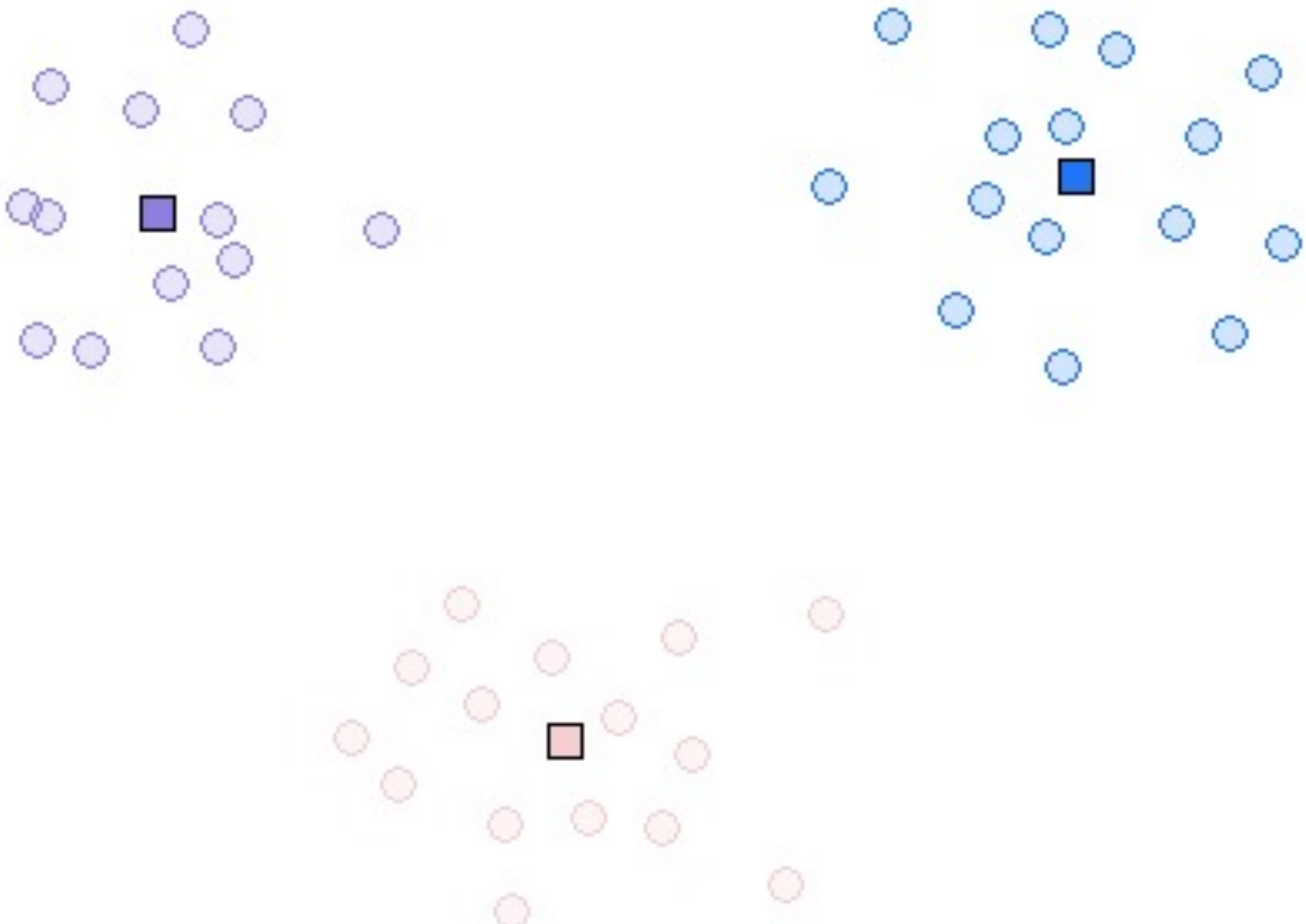
Again duality: L2-loss, MLE with Gaussian





Voronoi Tessellation





K-Means algorithm

- Pick K cluster centroids at random
- Alternate until convergence
  - Assign points to nearest centroid
  - Set centroid to the mean of the examples assigned to it

K-means objective

Loss to be minimized

$$J(\mu, r) = \sum_{i=1}^{n} \sum_{k=1}^{K} r_{ik} ||\mu_k - \mathbf{x}_i||_2^2$$

r_{ik} 1 iff point x_i in cluster *k*

μ_k centroid of cluster *k*

Reconstruction error of approximating every x_i by the center of the cluster it is in

K-means algorithm

$$J(\mu, r) = \sum_{i=1}^{n} \sum_{k=1}^{K} r_{ik} ||\mu_k - \mathbf{x}_i||_2^2$$

Assign point *i* to cluster *k*

$$\arg\min_r J(\mu, r) \rightarrow r_{ik} = \mathbf{1}(k = \arg\min_{k'} ||\mu_{k'} - \mathbf{x}_i||_2^2)$$

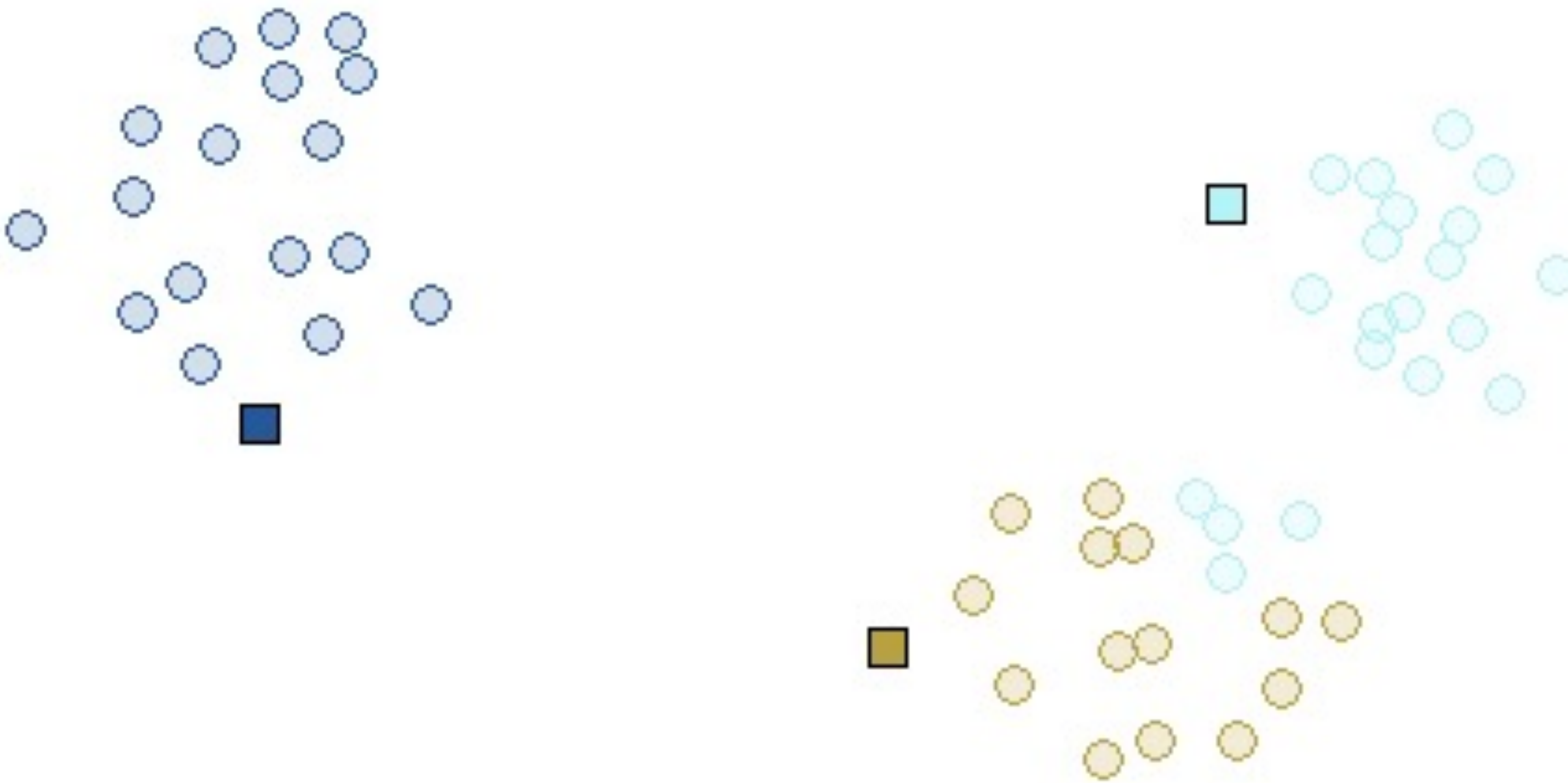
Compute centroid of cluster *k*

$$\arg\min_\mu J(\mu, r) \rightarrow \mu_k = \frac{\sum_i r_{ik}\mathbf{x}_i}{\sum_i r_{ik}}$$

Bad start

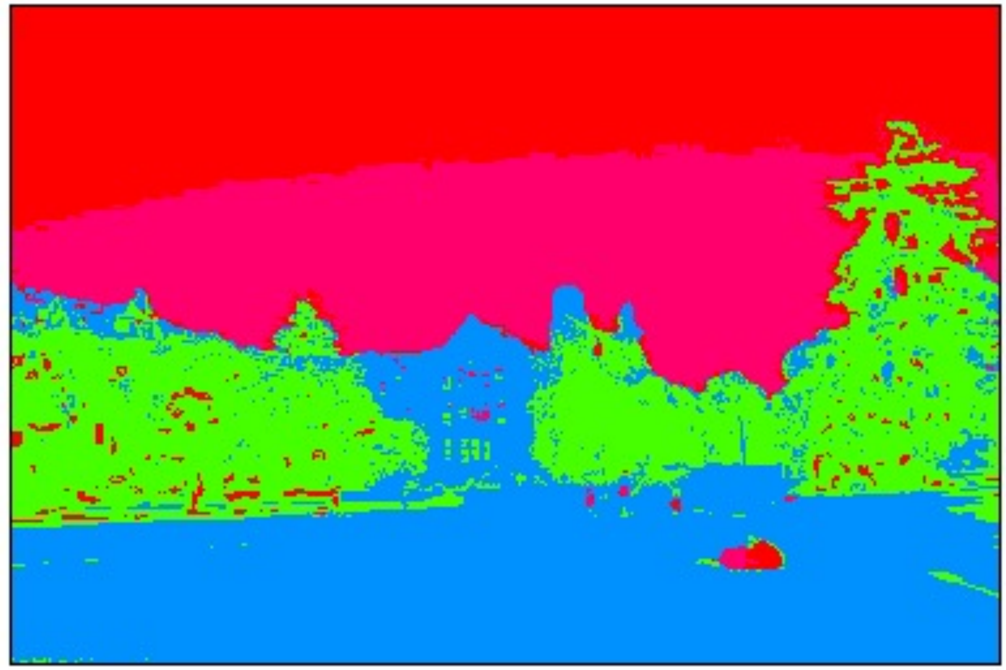
Bad start gives bad clusters

Better start



Better start gives better clusters

Example use - segmentation



What you should know

- **K-means uses alternating gradient descent**
  - estimates cluster membership and cluster centroids
  - minimizes reconstruction error
- **Usually initialized by selection of k random points, $\boldsymbol{x}_i$**
  - But can also pick points that are spread out
    - Kmeans++ "chooses centers at random from the data points, but weighs the data points according to their squared distance squared from the closest center already chosen."
- **Has a probabilistic interpretation (see the wiki for GMM/EM)**