# 1 Conceptual Review of K-means and mixture modeling

1. What is the benefit of using k-means algorithm when solving a partitioning problem?

2. Recap the steps to k-means algorithm

   1: Fixing the cluster centers, assign points to nearest clusters
   2: Given the point assignments, re-estimate cluster centers
   Termination: No points change clusters in next iteration

3. What is K-medoids, and how is it different from k-means? Discuss their pros and cons.

4. What is a key difference between mixture modeling and k-means?

## 2 Expectation Maximization (EM) with Gaussian Mixture Models (GMM)

Let $z$ be a multinomial random latent variable with components $z_1, z_2, ..., z_k$, where each component takes on 0 or 1 *i.e.* $P(z_j = 1)$ is the probability that a point comes from gaussian distribution $j$.

Let $\lambda = \mu_1, \mu_2, ..., \mu_k, \Sigma_1, ..., \Sigma_k, \pi_1, ..., \pi_k$ where $\pi_j$ = $P(z_j{=}1)$.

The log likelihood $\ell(\lambda|x_1, x_2, ..., x_m) = \sum_{i=1}^m log P(x_i|\lambda) = \sum_{i=1}^m log \sum_{j=1}^k \pi_j \mathcal{N}(x_i|\mu_j, \Sigma_j)$.
[Refer to EM lecture slide 17 for breaking down $P(x_i|\lambda)$ ]

(a) E-step: Calculate the posterior probability $P(z_j = 1|x_i, \lambda)$ $\forall i, j$.

$P(z_j = 1|x_i, \lambda)$

$= \frac{p(x_i|z_j=1,\mu_j,\Sigma_j)p(z_j=1|\pi_j)}{p(x_i|\lambda)}$ [Bayes Rule]

$= \frac{\mathcal{N}(x_i|\mu_j,\Sigma_j)\pi_j}{\sum_{l=1}^k \pi_l \mathcal{N}(x_i|\mu_l,\Sigma_l)}$ [Marginalization for denominator]

(Note: In lecture, Pat removed the denominator and represented the proportional probability with the numerator)

(b) M-step: Apply MLE and update the parameters $\pi_j$, $\mu_j$, $\Sigma_j$ $\forall j$.

$\frac{\partial \ell}{\partial \mu_j} = \frac{\partial \ell}{\partial \mu_j} \sum_{i=1}^m log \sum_{l=1}^k \pi_l \mathcal{N}(x_i|\mu_l, \Sigma_l)$ [Log likelihood function]

$= \sum_{i=1}^m \frac{1}{\sum_{l=1}^k \pi_l \mathcal{N}(x_i|\mu_l,\Sigma_l)} \frac{\partial \ell}{\partial \mu_j} \sum_{l=1}^k \pi_l \mathcal{N}(x_i|\mu_l, \Sigma_l)$ [Differentiation rule: $\frac{\partial}{\partial x} ln(u(x)) = \frac{1}{u(x)} * u'(x)$]

$= \sum_{i=1}^m \frac{1}{\sum_{l=1}^k \pi_l \mathcal{N}(x_i|\mu_l,\Sigma_l)} \frac{\partial \ell}{\partial \mu_j} \pi_j \mathcal{N}(x_i|\mu_j, \Sigma_j)$ [Eliminating terms with no $u_j$]

$= \sum_{i=1}^m \frac{\mathcal{N}(x_i|\mu_j,\Sigma_j)\pi_j}{\sum_{l=1}^k \pi_l \mathcal{N}(x_i|\mu_l,\Sigma_l)} \frac{\partial \ell}{\partial \mu_j} \frac{(x_i-\mu_j)^2}{2\Sigma_j}$ [Exponential rule: $\frac{\partial}{\partial x} e^{u(x)} = e^{u(x)} * u'(x)$]

$= \sum_{i=1}^m P(z_j = 1|x_i, \lambda) \frac{\partial \ell}{\partial \mu_j} \frac{(x_i-\mu_j)^2}{2\Sigma_j}$ [Substitute from E-step]

$= \sum_{i=1}^m P(z_j = 1|x_i, \lambda) \Sigma_j^{-1}(x_i - \mu_j)$ [Derivative of log gaussian density function]

Setting this to 0, you get: $\mu_j = \frac{\sum_{i=1}^m P(z_j=1|x_i,\lambda)x_i}{\sum_{i=1}^m P(z_j=1|x_i,\lambda)}$

Similar calculation produces $\Sigma_j$ and $\pi_j$.