LECTURE 28 - EXPECTATION MAXIMIZATION FOR GAUSSIAN MIXTURE MODEL CLUSTERING CONTINUED

- Last class we introduced the Gaussian Mixture Model:
  - $p(x) = \sum_{k=1}^{K} \pi_k N(x|\mu_k, \Sigma_k)$ where $0 \leq \pi_k \leq 1$ and $\sum_k \pi_k = 1$
  - Suppose we are given $X = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ where each $\mathbf{x}_i$ is a sample from one of the K Gaussians in our mixture model. We want to estimate π_k, μ_k, Σ_k given X.
  - So, we want to maximize the following data likelihood:

$$\hat{\Theta} = argmax_{\Theta} \prod_{i=1}^{N} \sum_{k=1}^{K} \pi_k N(x|\mu_k, \Sigma_k) \tag{1}$$

where $\Theta = \{\pi_k, \mu_k, \Sigma_k\}_{k=1}^{K}$

- It is difficult to maximize! We should try to simpler version in which we add latent variables to simplify the problem (and apply EM).
- The hidden/latent/missing variable we added was the label of the Gaussian from which $\mathbf{x}_i$ was drawn

$$\mathbf{x}, z \sim f(\mathbf{x}, z|\theta) \tag{2}$$

- The complete data likelihood is then:

$$\begin{aligned} L^c &= \prod_{i=1}^{N} p(\mathbf{x}_i|z_i, \theta) p(z_i) && (3) \\ &= \prod_{i=1}^{N} N(\mathbf{x}_i|\mu_{z_i}, \theta_{z_i}) \pi_{z_i} && (4) \end{aligned}$$

- Since we do not know the z_i values, we do not just guess one value - we average over all possible values for z_i. In other words, we take the *expectation* of the complete likelihood with respect to z

$$
\begin{aligned}
Q(\Theta, \Theta^t) &= \mathbb{E}\left[\ln p(\mathbf{X}, \mathbf{z}|\Theta)|\mathbf{X}, \Theta^t\right] && (5)\\
&= \sum_{\mathbf{z}} p(\mathbf{z}|\mathbf{X}, \Theta^t) \ln p(\mathbf{X}, \mathbf{z}|\Theta) && (6)\\
&= \sum_{i=1}^{N} \sum_{z_i=1}^{K} p(z_i|\mathbf{x}_i, \Theta^t) \ln p(\mathbf{x}_i, z_i|\Theta) && (7)\\
&= \sum_{i=1}^{N} \sum_{z_i=1}^{K} p(z_i|\mathbf{x}_i, \Theta^t) \ln p(\mathbf{x}_i|z_i, \Theta) p(z_i) && (8)
\end{aligned}
$$

- Thus, to take the expectation, we need $p(z_i|\mathbf{x}_i, \Theta^t)$

$$
\begin{aligned}
p(z_i|\mathbf{x}_i, \Theta^t) &= \frac{\pi_{z_i}^t p_{z_i}(\mathbf{x}_i|\theta_{z_i}^t, z_i)}{p(\mathbf{x}_i|\Theta^t)} && (9)\\
&= \frac{\pi_{z_i}^t p_{z_i}(\mathbf{x}_i|\theta_{z_i}^t, z_i)}{\sum_{k=1}^{K} \pi_k^t p_k(\mathbf{x}_i|\theta_k^t, k)} && (10)
\end{aligned}
$$

- This completes the Expectation step in EM. Now, we must derive the update equations for the Maximization step. So, we need to maximize for π_k, Σ_k, μ_k.

1. Update Equation for mu_k

- For simplicity, let us assume that $\Sigma_k = \sigma_k^2 \mathbf{I}$

$$
\begin{aligned}
Q(\Theta, \Theta^t) &= \sum_{i=1}^{N} \sum_{z_i=1}^{K} p(z_i|\mathbf{x}_i, \Theta^t) \ln p(\mathbf{x}_i|z_i, \Theta) p(z_i) && (11)\\
&= \sum_{i=1}^{N} \sum_{k=1}^{K} p(z_i = k|\mathbf{x}_i, \Theta^t) \ln N(\mathbf{x}_i|\mu_k, \sigma_k^2)\pi_k && (12)\\
&= \sum_{i=1}^{N} \sum_{k=1}^{K} p(z_i = k|\mathbf{x}_i, \Theta^t) \ln N(\mathbf{x}_i|\mu_k, \sigma_k^2) + \ln \pi_k && (13)\\
&= \sum_{i=1}^{N} \sum_{k=1}^{K} p(z_i = k|\mathbf{x}_i, \Theta^t) \left(-\frac{d}{2} \ln \sigma_k^2 - \frac{1}{2\sigma_k^2} \left\|\mathbf{x}_i - \mu_k\right\|_2^2 + \ln \pi_k \right) && (14)
\end{aligned}
$$

$$\frac{\partial Q(\Theta,\Theta^t)}{\partial \mu_k} = \frac{\partial}{\partial \mu_k}\left[\sum_{i=1}^{N}\sum_{k=1}^{K} p(z_i=k|\mathbf{x}_i,\Theta^t)\left(-\frac{d}{2}\ln\sigma_k^2 - \frac{1}{2\sigma_k^2}\left\|\mathbf{x}_i-\mu_k\right\|_2^2 + \ln\pi_k\right)\right] = 0 \tag{15}$$

$$= \sum_{i=1}^{N} p(z_i=k|\mathbf{x}_i,\Theta^t)\left(\frac{1}{\sigma_k^2}(\mathbf{x}_i-\mu_k)\right) = 0 \tag{16}$$

$$= \sum_{i=1}^{N} p(z_i=k|\mathbf{x}_i,\Theta^t)\frac{\mathbf{x}_i}{\sigma_k^2} - \sum_{i=1}^{N} p(z_i=k|\mathbf{x}_i,\Theta^t)\frac{\mu_k}{\sigma_k^2} = 0 \tag{17}$$

$$= \sum_{i=1}^{N} p(z_i=k|\mathbf{x}_i,\Theta^t)\frac{\mathbf{x}_i}{\sigma_k^2} = \sum_{i=1}^{N} p(z_i=k|\mathbf{x}_i,\Theta^t)\frac{\mu_k}{\sigma_k^2} \tag{18}$$

$$= \sum_{i=1}^{N} p(z_i=k|\mathbf{x}_i,\Theta^t)\mathbf{x}_i = \sum_{i=1}^{N} p(z_i=k|\mathbf{x}_i,\Theta^t)\mu_k \tag{19}$$

$$\mu_k = \frac{\sum_{i=1}^{N} p(z_i=k|\mathbf{x}_i,\Theta^t)\mathbf{x}_i}{\sum_{i=1}^{N} p(z_i=k|\mathbf{x}_i,\Theta^t)} \tag{20}$$

2. Update Equation for σ_k^2

$$\frac{\partial Q(\Theta,\Theta^t)}{\partial \sigma_k^2} = \frac{\partial}{\partial \sigma_k^2}\left[\sum_{i=1}^{N}\sum_{k=1}^{K} p(z_i=k|\mathbf{x}_i,\Theta^t)\left(-\frac{d}{2}\ln\sigma_k^2 - \frac{1}{2\sigma_k^2}\left\|\mathbf{x}_i-\mu_k\right\|_2^2 + \ln\pi_k\right)\right] = 0 \tag{21}$$

$$= \sum_{i=1}^{N} p(z_i=k|\mathbf{x}_i,\Theta^t)\left[\left(-\frac{d}{2\sigma_k^2} + \frac{2}{2\left(\sigma_k^2\right)^2}\left\|\mathbf{x}_i-\mu_k\right\|_2^2\right)\right] = 0 \tag{22}$$

$$= \sum_{i=1}^{N} p(z_i=k|\mathbf{x}_i,\Theta^t)\frac{2d}{2\sigma_k^2} = \sum_{i=1}^{N} p(z_i=k|\mathbf{x}_i,\Theta^t)\frac{1}{\left(\sigma_k^2\right)^2}\left\|\mathbf{x}_i-\mu_k\right\|_2^2 \tag{23}$$

$$= \sum_{i=1}^{N} p(z_i=k|\mathbf{x}_i,\Theta^t)d = \sum_{i=1}^{N} p(z_i=k|\mathbf{x}_i,\Theta^t)\frac{1}{\sigma_k^2}\left\|\mathbf{x}_i-\mu_k\right\|_2^2 \tag{24}$$

$$= \sum_{i=1}^{N} p(z_i=k|\mathbf{x}_i,\Theta^t)d = \frac{1}{\sigma_k^2}\sum_{i=1}^{N} p(z_i=k|\mathbf{x}_i,\Theta^t)\left\|\mathbf{x}_i-\mu_k\right\|_2^2 \tag{25}$$

$$= \sigma_k^2 = \frac{\sum_{i=1}^{N} p(z_i=k|\mathbf{x}_i,\Theta^t)\left\|\mathbf{x}_i-\mu_k\right\|_2^2}{d\sum_{i=1}^{N} p(z_i=k|\mathbf{x}_i,\Theta^t)} \tag{26}$$

3. Update Equation for π_k

$$
\begin{aligned}
& \frac{\partial Q(\Theta, \Theta^t)}{\partial \pi_k} = \\
& \frac{\partial}{\partial \pi_k}\left[\sum_{i=1}^{N}\sum_{k=1}^{K} p(z_i = k|\mathbf{x}_i, \Theta^t)\left(-\frac{d}{2}\ln\sigma_k^2 - \frac{1}{2\sigma_k^2}\left\|\mathbf{x}_i - \mu_k\right\|_2^2 + \ln\pi_k\right) - \lambda\left(\sum_{k=1}^{K}\pi_k - 1\right)\right] = 0 \\
= \; & \sum_{i=1}^{N} p(z_i = k|\mathbf{x}_i, \Theta^t)\left(\frac{1}{\pi_k}\right) - \lambda = 0 \\
= \; & \left(\frac{1}{\pi_k}\right)\sum_{i=1}^{N} p(z_i = k|\mathbf{x}_i, \Theta^t) - \lambda = 0 \\
= \; & \sum_{i=1}^{N} p(z_i = k|\mathbf{x}_i, \Theta^t) = \lambda\pi_k
\end{aligned} \tag{27}
$$

Since $\sum_k \pi_k = 1$, then:

$$
\begin{gathered}
\sum_{k=1}^{K} \frac{\sum_{i=1}^{N} p(z_i=k|\mathbf{x}_i,\Theta^t)}{\lambda} = \sum_{k=1}^{K}\pi_k = 1 \\
\lambda = \sum_{k=1}^{K}\sum_{i=1}^{N} p(z_i = k|\mathbf{x}_i, \Theta^t)
\end{gathered} \tag{28}
$$

So:

$$
\begin{aligned}
\pi_k &= \frac{\sum_{i=1}^{N} p(z_i = k|\mathbf{x}_i, \Theta^t)}{\sum_{k=1}^{K}\sum_{i=1}^{N} p(z_i = k|\mathbf{x}_i, \Theta^t)} && (29) \\
&= \frac{\sum_{i=1}^{N} p(z_i = k|\mathbf{x}_i, \Theta^t)}{\sum_{i=1}^{N}\sum_{k=1}^{K} p(z_i = k|\mathbf{x}_i, \Theta^t)} && (30) \\
&= \frac{\sum_{i=1}^{N} p(z_i = k|\mathbf{x}_i, \Theta^t)}{\sum_{i=1}^{N} 1} && (31) \\
&= \frac{\sum_{i=1}^{N} p(z_i = k|\mathbf{x}_i, \Theta^t)}{N} && (32)
\end{aligned}
$$

- We now have everything we need to implement the EM algorithm.
- Pseudo-code for the algorithm is:
  - Initialize all parameters
  - t = 1
  - While convergence not yet reached:
    * E-step: Compute $p(z_i = k|\mathbf{x}_i, \Theta^t)$ for every $\mathbf{x}_i$ and k
    * M-step:
      · Update μ_k for all k
      · Update σ_k^2 for all k
      · Update π_k for all k
    * t = t+1
    * Check convergence criteria
- *Does the EM algorithm find a global optima?*
- *Run EM Code*