Supplementary Notes

- Low-rank approximation
- Avoiding numerical issues

Evaluating density of multivariate Gaussian

- Given data $x \in \mathbb{R}^m$, the likelihood that it comes from a multivariate Gaussian density with mean vector $\mu \in \mathbb{R}^m$ and covariance matrix $\Sigma \in \mathbb{R}^{m \times m}$ is

$$\mathcal{N}(x; \mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^m \text{det}(\Sigma)}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}$$

- The most expensive part to compute this is to evaluate Σ^{-1}, which has a complexity $\mathcal{O}(m^3)$.
- Moreover, when Σ is rank-deficient, i.e., there are close-to-zero eigenvalues, computing Σ^{-1} will return NAN (you cannot invert the matrix)

- Now let's resolve the numerical issues and speedy up the computation by compute using "low-rank approximation"
- Compute eigendecomposition

$$\Sigma = U \Lambda U^T$$

where $\Lambda = \text{diag}\{\lambda_1, \ldots, \lambda_m\}$ and the eigenvalues are ordered

$$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m$$

- The rank-r approximation ($r < d$) of Σ is

$$\tilde{\Sigma} = \tilde{U} \tilde{\Lambda} \tilde{U}^T$$

where $\tilde{U}$ is a m-by-r matrix formed by the first r columns of U, $\tilde{\Lambda} = \text{diag}\{\lambda_1, \ldots, \lambda_r\}$.
- Typically we will choose r such that at least $\lambda_r \gg 0$

- Now compute transform of data and parameters

$$\tilde{x} = \tilde{U}^T x$$
$$\tilde{\mu} = \tilde{U}^T \mu$$

- Compute $\tilde{\Lambda}^{-1} = \text{diag}\{\lambda_1^{-1}, \ldots, \lambda_r^{-1}\}$
- Note that

$$\det(\Sigma) = \prod_{i=1}^{m} \lambda_i, \quad \det(\tilde{\Sigma}) = \prod_{i=1}^{r} \lambda_i$$

- Finally, the density calculated by replacing Σ with $\tilde{\Sigma}$ is:

$$\mathcal{N}(x; \mu, \Sigma) \approx \frac{1}{\sqrt{(2\pi)^m \prod_{i=1}^{r} \lambda_i}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{r} \frac{(\tilde{x}_i - \tilde{\mu}_i)^2}{\lambda_i}\right\}$$

  where $\tilde{x}_i$ and $\tilde{\mu}_i$ denote the ith entry of $\tilde{x}$ and $\tilde{\mu}$, respectively.
- Note: you can play with different r to have a good tradeoff between accuracy and numerical stability

- Note that above we have used the following basic identity from linear algebra

$$\tilde{\Sigma}^{-1} = \tilde{U}\tilde{\Lambda}^{-1}\tilde{U}^T$$

and

$$\begin{aligned}
&(x-\mu)^T\tilde{\Sigma}^{-1}(x-\mu) \\
=&(x-\mu)^T\tilde{U}\tilde{\Lambda}^{-1}\tilde{U}^T(x-\mu) \\
=&[\tilde{U}^T(x-\mu)]^T\tilde{\Lambda}^{-1}[\tilde{U}^T(x-\mu)] \\
=&[\tilde{x}-\tilde{\mu}]^T\tilde{\Lambda}^{-1}[\tilde{x}-\tilde{\mu}] \\
=&\sum_{i=1}^{r}\frac{(\tilde{x}_i-\tilde{\mu}_i)^2}{\lambda_i}
\end{aligned}$$

Avoiding numerical issues in GMM-EM

- Note that in evaluating E-step

$$\tau_k^i = \frac{\pi_k \mathcal{N}(x_i|\mu_k, \Sigma_k)}{\sum_{k'=1}^{K} \pi_{k'} \mathcal{N}(x_i|\mu_{k'}, \Sigma_{k'})}$$

  where the normal distributional density $\mathcal{N}(\cdot|\cdot, \cdot)$ appeared both in numerical and denominator

- Multivariate normal density

$$\mathcal{N}(X|\mu_k, \Sigma_k) := \frac{1}{|\Sigma|^{1/2}(2\pi)^{m/2}} \exp\left(-\frac{1}{2}(X-\mu)^T \Sigma^{-1}(X-\mu)\right)$$

  The term $(2\pi)^{m/2}$ can be very large when d is large

- So we can simplify the calculation without calculating $(2\pi)^{m/2}$ since it will be canceled out in the expression of τ_k^i
- Evaluate E-step based on low-rank approximation
  For $k = 1, \cdots, K$
  - Use low-rank approximation to compute, for each Gaussian component k

$$m_k = \sum_{i=1}^{r} \frac{(\tilde{x}_i - \tilde{\mu}_{i,k})^2}{\lambda_{i,k}}$$

$$D_k = \prod_{i=1}^{r} \lambda_{i,k}^{-1/2}$$

  - Compute

$$\hat{\tau}_k^i = \pi_k D_k \exp\left(-\frac{1}{2} m_k\right)$$

  Normalize

$$C = \sum_{k=1}^{K} \hat{\tau}_k^i$$

$$\tau_k^i = \hat{\tau}_k^i / C$$