CSE 326/426 (Fall 2022) Homework 6

Due on 11:55pm, Thu, Oct 21, 2022

Grading: All questions have the same points (25 each). Only questions with $*$ will be graded (undergraduates: do not need to answer the graduate-only question).

Submitting: Only electronic submissions in PDF format on Coursesite are accepted. Name your file as

<Your LIN>_HW6.pdf

A few ways to create a single PDF file. i) Use Microsoft Word and insert your writing or any figure taken, scanned, or plotted. Then choose "File→Print" in the main menu and you will find an option for outputting the file to a PDF file. ii) Use Latex to write your solution and include any figures. iii) Use the online Google Doc as an alternative of Word. It has sufficient features to combine multiple images and texts. Exporting to PDF files is similar as with Word.

Please DO NOT compress the PDF file as that will slow down the processing of your submission and the grading.

Questions:

1. Assume the data to be clustered are of n dimensions. In the distortion measure of the K-means objective function, count the number of parameters that you need to optimize. One scalar variable is counted as one parameter.

2. $*$ K-means on the real line. Given the unlabeled training data $x^{(1)} = 0$, $x^{(2)} = 1$, $x^{(3)} = -2$, find the two centers μ_1 and μ_2 on the real line and the corresponding assignment r_{ik} of the training data to the two clusters, so that the distortion measure $J(\mu_1, \mu_2) = \sum_{i=1}^{3} \sum_{k=1}^{2} r_{ik}(x^{(i)} - \mu_k)^2$ is minimized. What's the smallest J value?

   [*Hints: there are only a finite number of possible assignments with different* μ_k*, and you can exhaust them and select the optimal assignment. First assume the mean of the centers* $\bar{\mu} = \frac{1}{2}(\mu_1 + \mu_2)$ *that separates the three points into two sets. Using this threshold, you can express* J *as a quadratic function of* μ_1 *and* μ_2 *and then find* μ_1 *and* μ_2 *to minimize* J*.*]

3. $*$ The input vector $\mathbf{x}$ is generated from a Gaussian $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \Sigma_k)$ with probability ϕ_k, so that $\Pr(\mathbf{x}) = \sum_{k=1}^{K} \phi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \Sigma_k)$. Let the one-hot vector $\mathbf{z} \in \{0,1\}^K$, $\sum_k z_k = 1$ indicate which cluster $\mathbf{x}$ belongs to, and $\Pr(z_k = 1) = \phi_k$. Prove that

$$\sum_{k=1}^{K} \phi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \Sigma_k) = \sum_{\mathbf{z}} \prod_{k=1}^{K} \phi_k^{z_k} \prod_{k=1}^{K} \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \Sigma_k)^{z_k}. \tag{1}$$

4. Given the one-hot vector $\mathbf{z} \in \{0,1\}^K$ so that any distribution of $\mathbf{z}$ is a discrete distribution over the K values of $\mathbf{z}$. Let $\mathbf{x}$ be a random vector with distribution $\Pr(\mathbf{x})$. The joint distribution $\Pr(\mathbf{x}, \mathbf{z})$ and conditional distribution $\Pr(\mathbf{z}|\mathbf{x})$ can be defined accordingly so that $\Pr(\mathbf{x}, \mathbf{z}) = \Pr(\mathbf{x})\Pr(\mathbf{z}|\mathbf{x})$. Prove that for any distribution $q(\mathbf{z})$,

$$\ln \Pr(\mathbf{x}) = \sum_{\mathbf{z}} q(\mathbf{z}) \ln \frac{\Pr(\mathbf{x}, \mathbf{z})}{q(\mathbf{z})} - \sum_{\mathbf{z}} q(\mathbf{z}) \frac{\Pr(\mathbf{z}|\mathbf{x})}{q(\mathbf{z})}. \tag{2}$$

5. $*$ (Graduate only) For the Gaussian mixture model (GMM) with K Gaussian components, let the incomplete-data log-likelihood of an observed data point $\mathbf{x} \in \mathbb{R}^n$ be

$$\log \Pr(\mathbf{x}|\pi, \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K, \Sigma_1, \ldots, \Sigma_K) = \log \left\{ \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \Sigma_k) \right\}, \tag{3}$$

where $\boldsymbol{\mu}_k, \Sigma_k$ are the mean and covariance of the k-th Gaussian, which has density $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \Sigma_k)$. To estimate the mean vector during the EM algorithm of GMM, you will need to find the partial derivative of the above log-likelihood with respect to $\boldsymbol{\mu}_k$. Prove that the partial derivative is

$$\frac{\pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_j, \Sigma_j)} \Sigma_k^{-1}(\mathbf{x} - \boldsymbol{\mu}_k). \tag{4}$$

[*Hints: you need to write down the full expression of the multi-variate Gaussian distribution, and take the gradient of it with respect to* $\boldsymbol{\mu}_k$. *Chain rule and the matrix calculus will be used too.*]