Introduction of Machine Learning

Ex4: Gaussian Mixture Models and Expectation-Maximization

TA: Oscar Wu, oscarwu217@alum.ccu.edu.tw
Lecturer: Prof. Peggy Lu, peggylu@cs.ccu.edu.tw

Problem 1

Suppose that $y \sim \mathcal{N}(\mu, \Sigma)$ and $z \sim \mathcal{N}(\hat{\mu}, \hat{\Sigma})$ are independent Gaussian distributed random variables, where $\mu, \hat{\mu} \in \mathbb{R}^n$, and $\Sigma, \hat{\Sigma} \in \mathbb{S}^n$. Then, prove that their sum is also Gaussian:

$$y + z \sim \mathcal{N}(\mu + \hat{\mu}, \Sigma + \hat{\Sigma}).$$

Problem 2

Show that for a mixture of Gaussians in which all components have a covariance ϵI, when $\epsilon \to 0$, maximizing the expected complete-data log likelihood is equivalent to minimizing the distortion of the K-means algorithm.

Problem 3

Consider that the regressors $x^{(i)}$, where $i = 1, \dots, D$, influence the output random variable y. The distribution of y is:

$$p(y_i = y|\boldsymbol{\theta}) = p\frac{\lambda_{1i}^y \exp(-\lambda_{1i})}{y!} + (1-p)\frac{\lambda_{2i}^y \exp(-\lambda_{2i})}{y!},$$

where $p(y_i = y|\boldsymbol{\theta})$ is a mixture of two Poisson distributions,

$$\lambda_{1i} = \exp\Big(\beta_1 x^{(i)}\Big), \quad \lambda_{2i} = \exp\Big(\beta_2 x^{(i)}\Big)$$

and

$$\boldsymbol{\theta} = (\beta_1, \beta_2, p)^T.$$

(a) Which are the minimum conditions for the above model to be identifiable?

(b) Explain (giving details of $Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{\text{old}})$ and the EM steps) how the EM algorithm can be used to obtain an estimator of $\boldsymbol{\theta}$.

(c) Derive the derivative of $Q(\boldsymbol{\theta}, \boldsymbol{\theta}_{\text{old}})$ and explain how the derivative may be useful in the maximization stage of the EM algorithm.

(d) Given an initial value, will the EM algorithm always find the maximum of the likelihood?

(e) Explain how one can check whether the parameter which maximizes the EM algorithm maximizes the likelihood.