Department of Computer Science and Technology

Machine Learning

Homework 3

Sahand Sabour

2020280401

1 Clustering: Mixture of Multinomials

1.1 MLE for multinomial

The likelihood function for this multinomial distribution is given as

$$P(x|\mu) = \frac{n!}{\prod_i x_i!} \prod_i \mu_i^{x_i}, \quad i = 1, ..., d \tag{1}$$

Taking log from both side of the above equation gives the log-likelihood function

$$\mathcal{L}(\mu) = log(P(x|\mu)) = log(n!) - log(\prod_i x_i!) + log(\prod_i \mu_i^{x_i}) \tag{2}$$

This can be considered a Lagrange problem with the constraint $\sum_i \mu_i = 1$. Hence, the Lagrangian equation can be formulated as

$$\mathcal{L}(\mu) = log(n!) - log(\prod_i x_i!) + log(\prod_i \mu_i^{x_i}) - \lambda(\sum_i \mu_i - 1) \tag{3}$$

where λ is Lagrangian multiplier, giving

$$\mathcal{L}(\mu) = log(n!) - \sum_i log(x_i!) + \sum_i x_i log(\mu_i) - \lambda(\sum_i \mu_i - 1) \tag{4}$$

Taking the derivative of the equation with respect to μ_i and setting it to 0 gives

$$\frac{\partial \mathcal{L}}{\partial \mu_i} = \frac{\sum_i x_i}{\sum_i \mu_i} - \lambda = 0 \tag{5}$$

Hence, we get that

$$\lambda = \frac{\sum_i x_i}{\sum_i \mu_i} = \frac{n}{1} = n \tag{6}$$

Accordingly, we could derive the maximum-likelihood estimator μ_i as

$$\mu_i = \frac{x_i}{\lambda} = \frac{x_i}{n}, \quad i = 1, ..., d \tag{7}$$

1.2 EM for mixture of multinomials

2 PCA

2.1 Minimum Error Formulation

3 Deep Generative Models: Class-conditioned VAE