Assignment 1: EM for Categorical Data Advanced Signal Processing

Daniel Barrejón Moreno

21 de noviembre de 2018

Formulation of the problem

Mixture model of categorical

$$p(\mathbf{x}|\boldsymbol{\theta}) = \sum_{k=1}^{K} \pi_k \prod_{j=1}^{D} \mathrm{Cat}(x_j|\boldsymbol{\theta}_k) \tag{1}$$

Incomplete data likelihood

$$p(\mathbf{X}, \boldsymbol{\theta}) = \prod_{i=1}^{N} \sum_{k=1}^{K} \pi_k \prod_{j=1}^{D} \mathrm{Cat}(x_{ij}|\boldsymbol{\theta}) \tag{2}$$

PDF x given z

$$p(\mathbf{x}_i|z_i, \boldsymbol{\theta}) = \prod_{k=1}^{K} \prod_{j=1}^{D} \mathrm{Cat}(x_{ij}|\boldsymbol{\theta}_k)^{\mathbb{I}(z_i=k)} \tag{3}$$

Marginal for z

$$p(z_i|\boldsymbol{\theta}) = \prod_{k=1}^{K} \pi_k^{\mathbb{I}(z_i=k)} \tag{4}$$

Marginal for x

$$p(\mathbf{x}_i) = \prod_{j=1}^{D} \mathrm{Cat}(x_{ij}|\boldsymbol{\theta}) = \prod_{j=1}^{D} \prod_{m=1}^{I} \theta_m^{\mathbb{I}(x_{i,j}=m)} = \prod_{m=1}^{I} \prod_{j=1}^{D} \theta_m^{\mathbb{I}(x_{i,j}=m)} \tag{5}$$

$$= \prod_{m=1}^{I} \theta_m^{\sum_{j=1}^{D} \mathbb{I}(x_{i,j}=m)} = \prod_{m=1}^{I} \theta_m^{\mu_{i,m}}, \tag{6}$$

where

$$\mu_{i,m} = \sum_{j=1}^{D} \mathbb{I}(x_{i,j} = m) \tag{7}$$

1. Log-likelihood l_c

Complete data log-likelihood

$$l_c(\boldsymbol{\theta}) = \ln p(\mathcal{D}, \mathcal{Z}|\boldsymbol{\theta}) = \sum_{i=1}^{N} \log(p(\mathbf{x}_i|\mathbf{z}_i, \boldsymbol{\theta})p(\mathbf{z}_i|\boldsymbol{\theta})) = \sum_{i=1}^{N} \log\left(\prod_{k=1}^{K}(\pi_k \prod_{j=1}^{D} \mathrm{Cat}(x_{ij}|\boldsymbol{\theta}_k))^{\mathbb{I}(z_i=k)}\right) \tag{8}$$

$$= \sum_{i=1}^{N}\sum_{k=1}^{K} \mathbb{I}(z_i = k) \log\left(\pi_k \prod_{j=1}^{D} \mathrm{Cat}(x_{ij}|\boldsymbol{\theta}_k)\right) \tag{9}$$

$$= \sum_{i=1}^{N}\sum_{k=1}^{K} \mathbb{I}(z_i = k) \log \pi_k + \sum_{i=1}^{N}\sum_{k=1}^{K} \mathbb{I}(z_i = k) \log \prod_{m=1}^{I} \theta_m^{\mu_{i,m}} \tag{10}$$

$$= \sum_{i=1}^{N}\sum_{k=1}^{K} \mathbb{I}(z_i = k) \log \pi_k + \sum_{i=1}^{N}\sum_{k=1}^{K} \mathbb{I}(z_i = k) \sum_{m=1}^{I} \mu_{i,m} \log \theta_m \tag{11}$$

2. 3. E-step: Q function

Expectation Step:

Taking the expectation with respect to the latent variables z we get the following expected complete data log-likelihood.

$$Q(\boldsymbol{\theta}^t, \boldsymbol{\theta}^{t-1}) = \mathbb{E}_Z\{l_c(\boldsymbol{\theta}|\mathcal{D}, \boldsymbol{\theta}^{t-1})\} = \mathbb{E}_Z\{\sum_{i=1}^{N}\sum_{k=1}^{K} \mathbb{I}(z_i = k) \log \pi_k + \sum_{i=1}^{N}\sum_{k=1}^{K} \mathbb{I}(z_i = k) \sum_{m=1}^{I} \mu_{i,m} \log \theta_m\} \tag{12}$$

$$= \sum_{i=1}^{N}\sum_{k=1}^{K} \mathbb{E}_Z\{\mathbb{I}(z_i = k)\} \log \pi_k + \sum_{i=1}^{N}\sum_{k=1}^{K} \mathbb{E}_Z\{\mathbb{I}(z_i = k)\} \sum_{m=1}^{I} \mu_{i,m} \log \theta_m \tag{13}$$

$$= \sum_{i=1}^{N}\sum_{k=1}^{K} r_{i,k} \log \pi_k + \sum_{i=1}^{N}\sum_{k=1}^{K} r_{i,k} \sum_{m=1}^{I} \mu_{i,m} \log \theta_m, \tag{14}$$

where

$$r_{i,k} = \mathbb{E}_Z\{\mathbb{I}(z_i = k)\} = p(z_i = k|\mathbf{x}_i, \boldsymbol{\theta}^{t-1}) = \frac{p(z_i = k, \mathbf{x}_i|\boldsymbol{\theta}^{t-1})}{p(\mathbf{x}_i|\boldsymbol{\theta}^{t-1})} = \frac{p(z_i = k, \mathbf{x}_i|\boldsymbol{\theta}^{t-1})}{\sum_{k\prime}^{K} p(z_i = k\prime, \mathbf{x}_i|\boldsymbol{\theta}^{t-1})} \tag{15}$$

$$= \frac{\pi_k p(x_i|\boldsymbol{\theta}_k)}{\sum_{k\prime}^{K} \pi_{k\prime} p(x_i|\boldsymbol{\theta}_{k\prime})} = \frac{\pi_k \prod_{j=1}^{D} \mathrm{Cat}(x_{i,j}|\boldsymbol{\theta}_k)}{\sum_{k\prime}^{K} \pi_{k\prime} \prod_{j=1}^{D} \mathrm{Cat}(x_{i,j}|\boldsymbol{\theta}_k} \tag{16}$$

3. M-step: Maximization

3.1. Maximization of π_k

$$\hat{\pi_k} = \frac{\sum_{i=1}^{N} r_{i,k}}{N} = \frac{N_k}{N}, \tag{17}$$

where

$$N_k = \sum_{i=1}^{N} r_{i,k} \tag{18}$$

3.2. Maximization of $\boldsymbol{\theta}_k$

$$\hat{\boldsymbol{\theta}}_{k,m} = \frac{\sum_{i=1}^{N} r_{i,k}\mu_{i,m}}{\sum_{m=1}^{I} \sum_{i=1}^{N} r_{i,k}\mu_{i,m}} \tag{19}$$