

Assignment 2: HMM for Categorical Data Sequences

Daniel Barrejón Moreno

14 de enero de 2019

1. Complete data log likelihood

We have the following marginal likelihood

$$p(\mathbf{y}_t | s_t = k, \mathbf{B}) = \prod_{j=1}^{D_t} \text{Cat}(y_{tj} | \mathbf{b}_k). \quad (1)$$

We need to find the complete data log likelihood for the joint distribution of S and Y for N sequences. First of all, the likelihood will look as follows

$$p(S, Y | \boldsymbol{\theta}) = p(Y | S, \boldsymbol{\theta}) p(S | \boldsymbol{\theta}) = \prod_{n=1}^N \left(p(s_1^n | \boldsymbol{\pi}) \prod_{t=2}^{T_n} p(s_t^n | s_{t-1}^n | \mathbf{A}) \right) \left(\prod_{t=1}^{T_n} p(\mathbf{y}_t^n | s_t^n | \mathbf{B}) \right), \quad (2)$$

and taking logarithms we get,

$$\log p(S, Y | \boldsymbol{\theta}) = \log \prod_{n=1}^N \left(p(s_1^n | \boldsymbol{\pi}) \prod_{t=2}^{T_n} p(s_t^n | s_{t-1}^n, \mathbf{A}) \right) \left(\prod_{t=1}^{T_n} p(\mathbf{y}_t^n | s_t^n, \mathbf{B}) \right) \quad (3)$$

$$= \quad (4)$$

Mixture model of categorical

$$p(\mathbf{x}|\boldsymbol{\theta}) = \sum_{k=1}^K \pi_k \prod_{j=1}^D \text{Cat}(x_j|\boldsymbol{\theta}_k) \quad (5)$$

Incomplete data likelihood

$$p(\mathbf{X}, \boldsymbol{\theta}) = \prod_{i=1}^N \sum_{k=1}^K \pi_k \prod_{j=1}^D \text{Cat}(x_{ij}|\boldsymbol{\theta}) \quad (6)$$

PDF \mathbf{x} given \mathbf{z}

$$p(\mathbf{x}_i|z_i, \boldsymbol{\theta}) = \prod_{k=1}^K \prod_{j=1}^D \text{Cat}(x_{ij}|\boldsymbol{\theta}_k)^{\mathbb{I}(z_i=k)} \quad (7)$$

Marginal for \mathbf{z}

$$p(z_i|\boldsymbol{\theta}) = \prod_{k=1}^K \pi_k^{\mathbb{I}(z_i=k)} \quad (8)$$

Marginal for \mathbf{x}

$$p(\mathbf{x}_i) = \prod_{j=1}^D \text{Cat}(x_{ij}|\boldsymbol{\theta}) = \prod_{j=1}^D \prod_{m=1}^I \theta_m^{\mathbb{I}(x_{i,j}=m)} = \prod_{m=1}^I \prod_{j=1}^D \theta_m^{\mathbb{I}(x_{i,j}=m)} \quad (9)$$

$$= \prod_{m=1}^I \theta_m^{\sum_{j=1}^D \mathbb{I}(x_{i,j}=m)} = \prod_{m=1}^I \theta_m^{\mu_{i,m}}, \quad (10)$$

where

$$\mu_{i,m} = \sum_{j=1}^D \mathbb{I}(x_{i,j} = m) \quad (11)$$

2. Log-likelihood l_c

Complete data log-likelihood

$$l_c(\boldsymbol{\theta}) = \ln p(\mathcal{D}, \mathcal{Z}|\boldsymbol{\theta}) = \sum_{i=1}^N \log(p(\mathbf{x}_i|\mathbf{z}_i, \boldsymbol{\theta})p(\mathbf{z}_i|\boldsymbol{\theta})) = \sum_{i=1}^N \log \left(\prod_{k=1}^K (\pi_k \prod_{j=1}^D \text{Cat}(x_{ij}|\boldsymbol{\theta}_k))^{\mathbb{I}(z_i=k)} \right) \quad (12)$$

$$= \sum_{i=1}^N \sum_{k=1}^K \mathbb{I}(z_i = k) \log \left(\pi_k \prod_{j=1}^D \text{Cat}(x_{ij}|\boldsymbol{\theta}_k) \right) \quad (13)$$

$$= \sum_{i=1}^N \sum_{k=1}^K \mathbb{I}(z_i = k) \log \pi_k + \sum_{i=1}^N \sum_{k=1}^K \mathbb{I}(z_i = k) \log \prod_{m=1}^I \theta_{k,m}^{\mu_{i,m}} \quad (14)$$

$$= \sum_{i=1}^N \sum_{k=1}^K \mathbb{I}(z_i = k) \log \pi_k + \sum_{i=1}^N \sum_{k=1}^K \mathbb{I}(z_i = k) \sum_{m=1}^I \mu_{i,m} \log \theta_{k,m} \quad (15)$$

3. 3. E-step: Q function

Expectation Step: xn

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

$$Q(\boldsymbol{\theta}^t, \boldsymbol{\theta}^{t-1}) = \mathbb{E}_Z\{l_c(\boldsymbol{\theta}|\mathcal{D}, \boldsymbol{\theta}^{t-1})\} = \mathbb{E}_Z\left\{\sum_{i=1}^N \sum_{k=1}^K \mathbb{I}(z_i = k) \log \pi_k + \sum_{i=1}^N \sum_{k=1}^K \mathbb{I}(z_i = k) \sum_{m=1}^I \mu_{i,m} \log \theta_{k,m}\right\} \quad (16)$$

$$= \sum_{i=1}^N \sum_{k=1}^K \mathbb{E}_Z\{\mathbb{I}(z_i = k)\} \log \pi_k + \sum_{i=1}^N \sum_{k=1}^K \mathbb{E}_Z\{\mathbb{I}(z_i = k)\} \sum_{m=1}^I \mu_{i,m} \log \theta_{k,m} \quad (17)$$

$$= \sum_{i=1}^N \sum_{k=1}^K r_{i,k} \log \pi_k + \sum_{i=1}^N \sum_{k=1}^K r_{i,k} \sum_{m=1}^I \mu_{i,m} \log \theta_{k,m}, \quad (18)$$

where

$$r_{i,k} = \mathbb{E}_Z\{\mathbb{I}(z_i = k)\} = p(z_i = k | \mathbf{x}_i, \boldsymbol{\theta}^{t-1}) = \frac{p(z_i = k, \mathbf{x}_i | \boldsymbol{\theta}^{t-1})}{p(\mathbf{x}_i | \boldsymbol{\theta}^{t-1})} = \frac{p(z_i = k, \mathbf{x}_i | \boldsymbol{\theta}^{t-1})}{\sum_{k'}^K p(z_i = k', \mathbf{x}_i | \boldsymbol{\theta}^{t-1})} \quad (19)$$

$$= \frac{\pi_k p(x_i | \boldsymbol{\theta}_k)}{\sum_{k'}^K \pi_{k'} p(x_i | \boldsymbol{\theta}_{k'})} = \frac{\pi_k \prod_{j=1}^D \text{Cat}(x_{i,j} | \boldsymbol{\theta}_k)}{\sum_{k'}^K \pi_{k'} \prod_{j=1}^D \text{Cat}(x_{i,j} | \boldsymbol{\theta}_{k'})} \quad (20)$$

4. M-step: Maximization

4.1. Maximization of π_k

$$\hat{\pi}_k = \frac{\sum_{i=1}^N r_{i,k}}{N} = \frac{N_k}{N}, \quad (21)$$

where

$$N_k = \sum_{i=1}^N r_{i,k} \quad (22)$$

4.2. Maximization of θ_k

$$\hat{\theta}_{k,m} = \frac{\sum_{i=1}^N r_{i,k} \mu_{i,m}}{\sum_{m=1}^I \sum_{i=1}^N r_{i,k} \mu_{i,m}} \quad (23)$$