ECE 544NA Fall 2016 Assignment 3

Jifu Zhao Date: 10/20/2016

I. Pencil-and-Paper

1. E-Step

Since that $p_{X|\Theta(x_i|\theta)}$ is described as:

$$p_{X|\Theta(x_i|\theta)} = \sum_{h=1}^{m} w_h p_{V|H,\Theta}(x_i, h, \theta) \tag{1}$$

Now, suppose the hidden variable is y, where y could be 1, 2, $\cdots$, m. Then, we can calculate the expectation of the log-likelihood $E[log\mathcal{L}(\Theta)]$ as:

$$E[log\mathcal{L}(\Theta)] = \sum_{k=1}^{m}[\sum_{i=1}^{n} log(\sum_{h=1}^{m} w_h p_{V|H,\Theta}(x_i, h, \theta|y = k))] \cdot \gamma_i(k) \tag{2}$$

where $\gamma_i(k)$ is the posterior probability for $y = k$, and $\gamma_i(k)$ is defined as:

$$\gamma_i(k) = \frac{w_k \mathcal{N}(x_i; \mu_k, \Sigma_k)}{\sum_l w_l \mathcal{N}(x_i; \mu_l, \Sigma_l)} \tag{3}$$

Also, notice that

$$p_{V|H,\Theta}(x_i, h, \theta|y = k)) = \begin{cases} \mathcal{N}(x_i; \mu_k, \Sigma_k) & \text{if } h = k \\ 0 & \text{if } h \neq k \end{cases} \tag{4}$$

So,

$$\begin{aligned} E[log\mathcal{L}(\Theta)] &= \sum_{i=1}^{n}\sum_{k=1}^{m} log(w_k p_{V|H,\Theta}(x_i, h = k, \theta)) \cdot \gamma_i(k) \\ &= \sum_{i=1}^{n}\sum_{k=1}^{m} log(w_k \mathcal{N}(x_i; \mu_k, \Sigma_k)) \cdot \frac{w_k \mathcal{N}(x_i; \mu_k, \Sigma_k)}{\sum_l w_l \mathcal{N}(x_i; \mu_l, \Sigma_l)} \\ &= \sum_{i=1}^{n}\sum_{k=1}^{m} log(w_k \mathcal{N}(x_i; \mu_k, \Sigma_k)) \cdot \gamma_i(k) \end{aligned} \tag{5}$$

So, in E-Step, the most important thing is to calculate $\gamma_i(k)$ for each i and k, where

$$\gamma_i(k) = \frac{w_k \mathcal{N}(x_i; \mu_k, \Sigma_k)}{\sum_l w_l \mathcal{N}(x_i; \mu_l, \Sigma_l)} \tag{6}$$

2. M-Step

From above equation (5), our goal is the maximize the expectation of log-likelihood $E[log\mathcal{L}(\Theta)]$.

First, consider w_k. With the constraint of $\sum_k w_k = 1$, we have:

$$\frac{\partial}{\partial w_k}[\sum_{i=1}^{n}\sum_{k=1}^{m} log(w_k\mathcal{N}(x_i;\mu_k,\Sigma_k))\cdot\gamma_i(k) + \lambda(\sum_k w_k - 1)] = 0 \tag{7}$$

So, we have:

$$\sum_{i=1}^{n}\frac{\gamma_i(k)}{w_k} + \lambda = 0 \tag{8}$$

Summing it over k from 1 to m and with the equation that $\sum_k \gamma_i(k) = 1$, we can have:

$$\lambda = n \tag{9}$$

So, we have:

$$w_k^{new} = \frac{1}{n}\sum_{i=1}^{n}\gamma_i(k) \tag{10}$$

Now, let's consider μ_k and Σ_k. Following the steps in *A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models*, we can simply get the following result:

$$\mu_k^{new} = \frac{\sum_{i=1}^{n} x_i\gamma_i(k)}{\sum_{i=1}^{n}\gamma_i(k)} \tag{11}$$

$$\Sigma_k^{new} = \frac{\sum_{i=1}^{n}(x_i - \mu_k^{new})\cdot(x_i - \mu_k^{new})^T\cdot\gamma_i(k)}{\sum_{i=1}^{n}\gamma_i(k)} \tag{12}$$

So, Equation (10), (11) and (12) are the main steps for M-Step.

Following Equation (6), (10), (11) and (12), we can iterate through E-Step and M-Step until reaching some stop criteria.

II. Code-from-Scratch

1. Methods

2. Results

III. TensorFlow

1. Methods

2. Results