Assignment #4

B05902120 / Yu-Ting, TSENG

May 16, 2019

EM Algorithm

Assume a sentence s is generated by sampling words from a multinomial mixture model with two components. One component is a unigram language model $p(w|\theta_1)$ capturing topic 1; the other is a unigram language model $p(w|\theta_2)$ capturing topic 2. λ is an unknown parameter denoting the probability of selecting $p(w|\theta_1)$ to generate a word in sentence s; $1-\lambda$ is the probability of selecting $p(w|\theta_2)$. Let $s = (w_1, w_2, ..., w_n)$, where $w_i(i = 1, 2, ..., n)$ is a word in vocabulary V. Given $p(w|\theta_1)$ and $p(w|\theta_2)$, please make use of the EM algorithm to compute the maximum likelihood estimate of λ

Give a formula to compute the log-likelihood of sentence s given the mixture model, i.e., the probability of observing all of the words in s being generated from the mixture model.

Mixture Model: $p(s|\theta) = \prod_{i=1}^{n} \lambda p(w_i|\theta_1) + (1-\lambda)p(w_i|\theta_2)$.
Likelihood Function: $L(\theta_1 \bigoplus \theta_2) = \sum_{i=1}^{n} \log[\lambda p(w_i|\theta_1) + (1-\lambda)p(w_i|\theta_2)]$.

How many binary hidden variables in total do we need for computing this maximum likelihood estimate using the EM algorithm? Why?

The number of binary hidden variables must equals to the number of the words, therefore the answer should be n.

Give the E-step and M-step updating formulas for estimating λ.

Let the hidden variable z_i represents which topic w_i it comes from. Moreover, we suppose all the words in the sentence only appear once.
E-step: Calculate the expectation value.
$Q(\lambda, \lambda^{(n)} = \sum_{i=1}^{n} p(z_i = 1|s, \lambda^{(n)}) \log(\lambda p(w_i|\theta_1)) + p(z_i = 0|s, \lambda^{(n)}) \log((1-\lambda)p(w_i|\theta_2))$
$p(z_i = 1|s, \lambda^{(n)}) = \frac{\lambda p(w_i|\theta_1)}{\lambda p(w_i|\theta_1) + (1-\lambda)p(w_i|\theta_2^n)}$
$p(z_i = 0|s, \lambda^{(n)}) = 1 - p(z_i = 1|s, \lambda^{(n)})$
M-step: Maximize the value of the Q function, and update to get the new λ.

$\frac{\partial Q(\lambda,\lambda^{(n)})}{\partial \lambda} = \sum_{i=1}^{n} p(z_i = 1|s, \lambda^{(n)}) \frac{p(w_i|\theta_1)}{\lambda p(w_i|\theta_1))} + \frac{-p(w_i|\theta_2)}{p(1-\lambda)p(w_i|\theta_2)} = 0$
$\sum_{i=1}^{n} \frac{p(z_i=1|s,\lambda^{(n)})-1}{1-\lambda} + \frac{p(z_i=1|s,\lambda^{(n)})}{\lambda} = 0$
$\lambda = \frac{p(z_i=1|s,\lambda^{(n)})}{n}$