Εργασία Εβδομάδας 7

Διονύσιος Κοτζαΐτσης*

1 Problem 1

We have n observations of i.i.d data. So:

$$p(v_1^n, ..., v_{i-1}^n, v_{i+1}^n, ..., v^N) = \int_{v_i} p(v) = \sum_h p(h) \left[\int_{v_i} p(v_i|h) \right] \prod_{j \neq i} p(v_j|h)$$

Knowing that $\int_{v_i} p(v_i|h) = 1$ from the laws of the continuous distributions, so:

$$p(v_1^n, ..., v_{i-1}^n, v_{i+1}^n, ..., v^N) = \int_{v_i} p(v) = \sum_h p(h) \prod_{j \neq i} p(v_j|h)$$

Which really means that we ignore the i-th element.

2 Problem 2

For the EM algorithm we need to use the energy term,

$$\sum_{n=1}^{N} < log[p(x^n|i)p(i)] >_{p^{old}(i|x^n)} =$$

$$\sum_{n=1}^{N} \sum_{i=1}^{H} p^{old}(i|x^n) \left[-\frac{1}{2}(x^n - \mu_i)^T \Sigma_i^{-1} (x^n - \mu_i) - \frac{1}{2} logdet(2\pi\Sigma_i) + logp(i) \right] \quad (1)$$

Now, we have to differentiate (1) with regards to $\mu, \Sigma, p(i)$:

2.1 optimal μ_i

We have to basically minimize:

$$\sum_{n=1}^{N} \sum_{i=1}^{H} p^{old}(i|x^n)(x^n - \mu_i)^T \Sigma_i^{-1} (x^n - \mu_i)$$

*Μεταπτυχιακός φοιτητής Τμ. Εφαρμοσμένης Πληροφορικής, Πανεπιστήμιο Μακεδονίας, aida22003@uom.gr

Differentiating with respect to μ and equalizing to zero, we have:

$$-2\sum_{n=1}^{N} p^{old}(i|x^n)\Sigma_i^{-1}(x^n - \mu_i) = 0$$

$$\mu_i = \frac{\sum_{n=1}^{N} p^{old}(i|x^n)x^n}{\sum_{n=1}^{N} p^{old}(i|x^n)}$$

Using a membership distribution, we have:

$$\mu_i = \sum_{n=1}^{N} p^{old}(n|i)x^n$$

2.2 p(i)

$$p^{new}(i) = \frac{1}{N}\sum_{n=1}^{N} p^{old}(i|x^n)$$

2.3 optimal Σ_i

This time we have to minimize the term :

$$\sum_{n=1}^{N}\sum_{i=1}^{H} p^{old}(i|x^n)\left[(x^n - \mu_i)^T\Sigma_i^{-1}(x^n - \mu_i) - logdet(\Sigma_i^{-1})\right]$$

Isolating the dependancy on Σ_i and using trace properties, we have:

$$trace\left(\Sigma_i^{-1}\sum_{n=1}^{N} p^{old}(i|x^n)(x^n - \mu_i)(x^n - \mu_i)^T\right) - logdet(\Sigma_i^{-1})\sum_{n=1}^{N} p^{old}(i|x^n)$$

Here, we have $\Sigma_i = diag(d_1^i, ...d_N^i)$,so :

$$trace\left(d_j^i\sum_{n=1}^{N} p^{old}(i|x^n)(x^n - \mu_i)(x^n - \mu_i)^T\right) - logd_j^i\sum_{n=1}^{N} p^{old}(i|x^n)$$

Differentiating with regards to d_j^i: By using $\frac{\partial}{\partial A}tr[AB] = B^T$, we have:

$$\sum_{n=1}^{N} p^{old}(i|x^n)(x^n - \mu_i)(x^n - \mu_i)^T\frac{1}{d_j^i} - logd_j^i\sum_{n=1}^{N} p^{old}(i|x^n)$$

Equating it to zero, we have:

$$d_j^i = \sum_{n=1}^{N} p^{old}(i|x^n)(x^n - \mu_i)(x^n - \mu_i)^T$$

3 Problem 3

We have:

$$p(n|i) = \frac{p(i|x^n)}{\sum_n p(i|x^n)} = \frac{p(x^n|i)p(i)}{\sum_n p(x^n|i)p(i)} = \frac{p(x^n|i)}{\sum_n p(x^n|i)}$$

If we replace with the Gaussian form, we have

$$p(n|i) = \frac{e^{-\frac{1}{2\sigma^2}(x^n - m_i)^2}}{\sum_n e^{-\frac{1}{2\sigma^2}(x^n - m_i)^2}}$$

If σ^2 becomes very small, the numerator will become very very small. But there will be some x^n so close to m that even thought the numerator is still small, is exponentially larger than all other x^n's

4 Problem 4

$$\frac{\partial}{\partial p(h)} L = \sum_n \frac{\partial}{\partial p(h)} \sum_{h'} p_{old}(h'|v') logp(h') - \lambda = \sum_n p_{old}(h|v^n) \frac{1}{p(h)} - \lambda$$

Equating to 0 to find the optimum, we have:

$$\sum_n p_{old}(h|v^n) \frac{1}{p(h)} - \lambda = 0$$

$$p(h) = \frac{1}{\lambda} \sum_n p_{old}(h|v^n)$$

5 Problem 5

Lets first find the MLE for covariance matrix:

$$LL = \sum_{i=1}^{n} \left(-\frac{1}{2} log(2\pi) - \frac{1}{2} log|\Sigma| - \frac{1}{2}(x^i - \mu)^T \Sigma^{-1} (x^i - \mu) \right)$$

$$LL = -\frac{n}{2} log(2\pi) - \frac{n}{2} log|\Sigma| - \frac{1}{2} \sum_{i=1}^{n} (x^n - \mu)^T \Sigma^{-1} (x^n - \mu)$$

Computing the derivative with regards to Σ, knowing that $\frac{\partial}{\partial A} log|A| = A^{-T}$, $\frac{\partial}{\partial A} tr[AB] = B^T$:

$$\frac{\partial LL}{\partial \Sigma} = \frac{n}{2} log|\Sigma^{-1}| - \frac{1}{2} \sum_{i=1}^{n} tr\left[(x^n - \mu)^T (x^n - \mu) \Sigma^{-1} \right]$$

$$\frac{\partial LL}{\partial \Sigma} = \frac{n}{2}\Sigma - \frac{1}{2}\sum_{i=1}^{n}(x^n - \mu)(x^n - \mu)^T$$

Since $\Sigma^T = \Sigma$.Equating to zero and solving for Σ, we have:

$$n\Sigma - \sum_{i=1}^{n}(x^n - \mu)(x^n - \mu)^T = 0$$

$$\Sigma = \frac{1}{n}\sum_{i=1}^{n}(x^n - \mu)(x^n - \mu)^T$$

, which is the MLE for Σ.
Where μ is the data mean, so for finite and bounded data the covariance remains finite.