Intelligent Systems: Reasoning and Recognition
Home task 6

Riyane SID-LAKHDAR

May 1, 2016

In this document, we will use the following notations:

- M: the number of samples(students)
- K: the number of sources (universities)
- D: the dimension of the samples (number of topics considered).
- ω_k: the event that a student comes from the university k.
- μ: the vector of dimension D representing the first moment of the samples: $\mu = \frac{\sum_{m=1}^{M} X_m}{M}$
- Σ: the matrix of size $(D * d)$ representing the second moment of the samples: $\Sigma[i][j] = \sigma_{i,j}^2 = E(X - E(X))$

1 Use a ratio of histograms to estimate the origin of a student from his grade

Suppose we have the histogram function h which for each vector $X \in (MARK_1, MARK_2)$ gives $h(x)$: number of students who had this marks.

Using the Bayes rule, we have:

$$P(\omega_k|X_m) = \frac{P(X_m|\omega_k) * P(\omega_k)}{P(X)} \quad = \frac{\frac{h_k(X_m)}{M_k} * \frac{M_k}{M}}{\frac{h(X)}{M}}$$

$$= \frac{h_k(X_m)}{h(X)}$$

The histogram need to store the number of occurrences of each couple of notes. As each note is an integer between 0 and 20, the dimension of the histogram is $Q = 20^2$.

To give reasonable results, we need to have at least enough samples to fill all the cells of the histogram (to identify the first case: equi-distribution). Thus the minimum number of samples must be proportional to the capacity of the histogram Q.

A more accurate estimation of this threshold value may be given using the root-mean-square sampling error between the histogram and the density function. Which gives a threshold value in the order of $\frac{Q}{M}$.

2 Normal density function to estimate the origin of a student from his grade

Once again we use the Bayes rule to define the probability for a given student m to come from a university k:

$$P(\omega_k|X_m) = \frac{P(X_m|\omega_k) * P(\omega_k)}{P(X)}$$

But this time we will compute each of this term using a normal density function:

$$\begin{aligned} P(X_m|\omega_k) &= N(X, \nu_k, \Sigma_k) \\ &= \frac{1}{(2\Pi)^{(D/2)} * det(\Sigma_k)^{(\frac{1}{2})}} * e^{(-\frac{1}{2}(X-\nu_k)^T * \sigma_k^{-1} * (X-\nu_k))} P(\omega_k) \quad = \frac{M_k}{M} \\ P(X) &= N(X, \nu, \Sigma) \\ &= \frac{1}{(2\Pi)^{(D/2)} * det(\Sigma_k)^{(\frac{1}{2})}} * e^{(-\frac{1}{2}(X-\nu_k)^T * \sigma_k^{-1} * (X-\nu_k))} \end{aligned}$$

Where μ and Σ are the first and second expectations of all the samples, and μ_k and Σ_k are the first and second expectations of the samples belonging to the class k.

Once the probabilities $P(k, m)$ have been computed for all the students m and all the origins k, we can compute the probability of error for each universiy k using the knowledge of the origin of each student:

$$\forall k \in [1, K] P_{error}(k) = 1 - \frac{FP(k)}{TP(k)} \tag{1}$$

Where

- FP(k): is the number of student m comming from the university k such as $\exists k_0 \in [1, K], P(k_0, m) > P(k, m)$

- TP(k): is the total number of student coming from the university k

3 "Expectation Maximization" to estimate the origin of a student from his grade

Let's only consider the note of the students for the 2 topics (suppose we don't know the origin of the students).
The objective is to find for each student m, and for each university k, the probability $P(k, m)$ that the student m comes from the university k.
Based on the **"Expectation Maximization"** algorithm, we compute this probability using a normal density function N:

$$
\begin{aligned}
\forall m \in [1, M] \\
\forall k \in [1, K] \\
P(k, m) &= N(X, \nu_k, \Sigma_k) \\
&= \frac{1}{(2\Pi)^{(D/2)} * det(\Sigma_k)^{(\frac{1}{2})}} * e^{(-\frac{1}{2}(X-\nu_k)^T * \sigma_k^{-1} * (X-\nu_k))}
\end{aligned}
$$

This operation is repeated, and for each iteration i, the value of ν_k and Σ_k are updated in order to minimize the Mahalanobis distance between the best sample (max probability) and the source:

$$
\begin{aligned}
\nu_k^{(i)} &= \frac{\sum_{m=1}^{M} h^{(i)}(k, m) * X}{\sum_{m=1}^{M} h^{(i)}(k, m)} \\
\Sigma_k^{(i)} &= \frac{\sum_{m=1}^{M} h^{(i)}(k, m) * (X_m - \mu_k) * (X_m - \mu_k)^T}{\sum_{m=1}^{M} h^{(i)}(k, m)}
\end{aligned} \tag{2}
$$

The processes is stopped after a fixed number of iterations, or using a quality metric Q^i (when Q^i becomes lower than a given threshold value).

The vector μ is initialized as a vector of 1 (or as the average of all the samples). The matrix Σ is initialized using the Identity matrix. No more accurate initialization may be used (cause we do further information about the order of the expected objective).