Chair of Computer Science 6
RWTH Aachen University
Prof. Dr.-Ing. H. Ney

Aachen, 22. 01. 2010

M. Nußbaum-Thom,
R. Schlüter

Automatic Speech Recognition

10. Exercise

Submission of the solutions: 29. 01. 2010 at the beginning of the lecture

---

Task 10.1 "Start–stop–detection" reconsidered.

In exercise 9 we introduced "start–stop–detection" using the squared error as local cost function.

In this task we reformulate the problem using a statistical approach. For each segment we assume a single gaussian distribution with mean μ_m and variance σ_m^2, $m = 1, 2, 3$. The desired segment boundaries t_{start} and t_{stop} for a given sequence of energies x_1^T are those time indices minimizing the negative log-likelihood calculated over the three segments.

In the following we distinguish two models:

(A) Pooled variance:
The variance σ_m^2, $m = 1, 2, 3$, is independent of the segment, i.e. we consider only one global variance σ^2 for all segments ($\sigma_m^2 = \sigma^2$, $m = 1, 2, 3$), while the mean μ_m still depends on segment m.

(B) Segment specific variances:
The variance σ_m^2 and the mean μ_m are both dependent on segment m.

(a) Prove that the log-likelihood calculated over the three segments equals (1)

$$\mathcal{LL}(x_1^T; \mu_m, \sigma_m^2, m = 1, 2, 3) = \sum_{m=1}^{3} \sum_{t=s_m}^{e_m} \log \mathcal{N}(x_t; \mu_m, \sigma_m^2),$$

where s_m is the starting time index and e_m the ending time index of segment m.
Remember, the log-likelihood of a sequence of observations x_1^N with respect to a single gaussian distribution with mean μ and variance σ^2 is given by

$$\mathcal{LL}(x_1^N; \mu, \sigma^2) = \sum_{n=1}^{N} \log \mathcal{N}(x_n; \mu, \sigma^2).$$

(b) Derive estimates for the unknown parameters of model (A) and (B), resp., by maximum-likelihood estimation. (2)
Hint: We assume that the sample sequence x_1^N is distributed according to a gaussian density with unknown mean μ and variance σ^2. The estimates of the unknown parameters derived by applying the maximum-likelihood approach are the empirical mean and variance of x_1^N

$$\hat{\mu} = \frac{1}{N}\sum_{n=1}^{N} x_n, \qquad \hat{\sigma}^2 = \frac{1}{N}\sum_{n=1}^{N}(x_n - \hat{\mu})^2.$$

(c) Prove that model (A) and the squared error criterion from exercise 4 yield the same segment boundaries. (2)

(d) Prove that for model (B) the cost, i.e. negative log-likelihood, for segment m, $m = 1, 2, 3$, with starting time index s_m and ending time index e_m is given by (3)

$$-\mathcal{LL}(x_{s_m}^{e_m}; \mu_m, \sigma_m^2) = \frac{(e_m - s_m + 1)}{2}\left[\log\left(\frac{1}{(e_m - s_m + 1)}\sum_{n=s_m}^{e_m}(x_n - \hat{\mu}_m)^2\right) + \log 2\pi + 1\right],$$

where $\hat{\mu}_m$ is the estimated mean.

(e) Implement an algorithm in C/C++ which computes the optimal segmentation according to the cost function derived in d). (4)

(f) Use your program from e) to segment the recording *probe1.ascii*. (4)
Plot *probe1.ascii* and draw in the derived segment boundaries. Attach the print out of your maximum likelihood estimates for the means and the variances. Compare the outcome with the results from exercise 4.

Use ANSI conformant C/C++ and send your source code by e-mail to

nussbaum@informatik.rwth-aachen.de.

The data file *probe1.ascii* is available at

http://www-i6.informatik.rwth-aachen.de/web/Teaching/Lectures/WS09_10/asr/index.html.