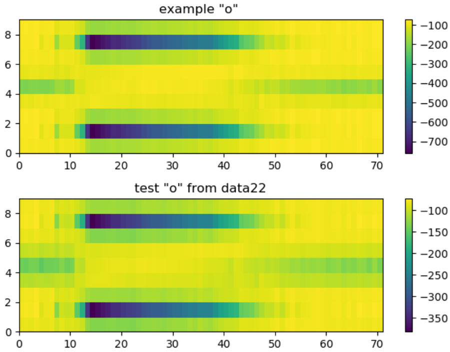
KTH ROYAL INSTITUTE
OF TECHNOLOGY

DT2119 Speech and Speaker Recognition Lab 2

Zijian Fan, Lingxi Xiong

April 30, 2019

5.1 Gaussian emission probabilities

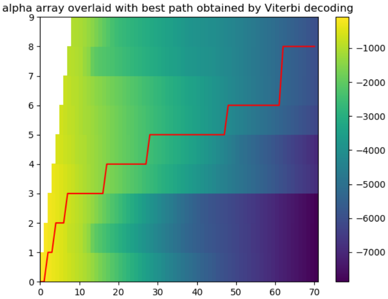


- isolated['o'] = ['sil'] + prondict['o'] + ['sil']

5.2 Forward Algorithm

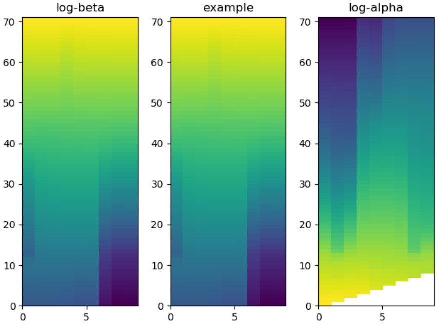
- Maximum likelihood accuracy
  trained on all speakers: 43/44
  trained on one speaker: 34/44
- Viterbi algorithm accuracy
  trained on all speakers: 44/44
  trained on one speaker: 34/44 (same mistakes as ML)
- Computational complexity
  Viterbi scoring: $O(n)$
  Forward scoring: $O(n^2)$

5.3 Viterbi Approximation



- $\alpha_n(j) = P(x_0, ..., x_n, z_n = s_j \mid \theta)$
- $logV_n(j) = max_{i=0}^{M-1}(logV_{n-1}(i) + loga_{ij}) + log\phi_j(x_n)$
- $B_n(j) = argmax_{i=0}^{M-1}(logV_{n-1}(i) + loga_{ij})$

5.4 Backward Algorithm

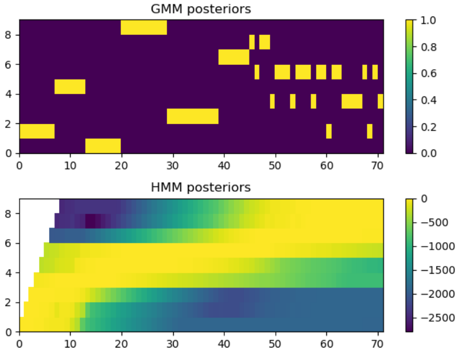


- $\beta_n(i) = P(x_{n+1}, \ldots, x_{N-1} \mid z_n = s_i, \theta)$

6.1 State posterior(Gamma) probabilities

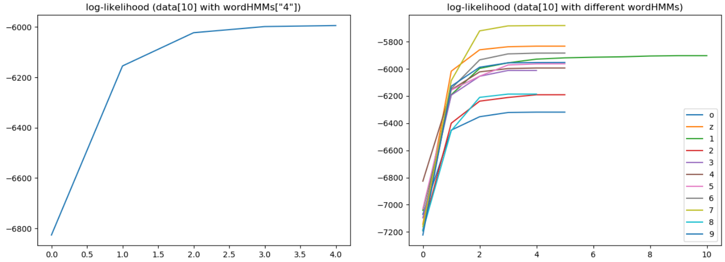
- HMM posteriors: $\gamma_n(i) = P(z_n = s_i \mid X, \theta)$
  For each time step the state posteriors sum to 1 (in linear domain).
  Summing along the time axis: 1.35, 2.10, 3.56, 9.74, 10.12, 20.53, 13.00, 1.21, 9.40
  Summing over both states and time steps = length of the observation(71)

6.1 State posterior(Gamma) probabilities



- HMM posteriors: $\gamma_n(i) = P(z_n = s_i \mid X, \theta)$
- GMM posteriors: $\gamma_n^{GMM}(i) = P(z_n = s_i \mid x_n, \phi)$

6.2 Retraining the emission probabilities distributions



Thank you for your Attention!