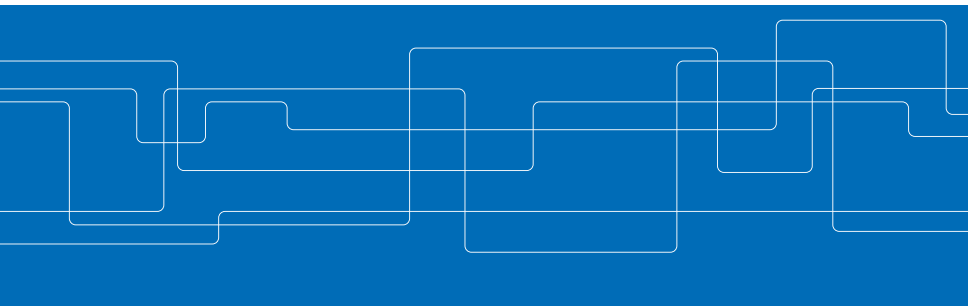




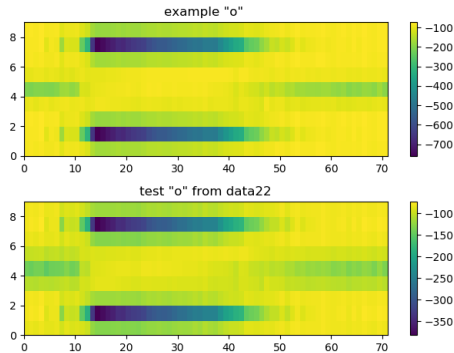
DT2119 Speech and Speaker Recognition Lab 2

Zijian Fan, Lingxi Xiong

May 1, 2019



5.1 Gaussian emission probabilities



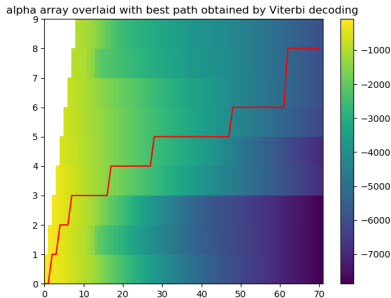
► $\text{isolated}['o'] = [\text{'sil'}] + \text{prondict}['o'] + [\text{'sil'}]$



5.2 Forward Algorithm

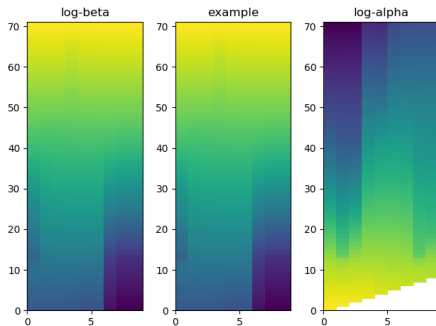
- ▶ Maximum likelihood accuracy
trained on all speakers: 43/44
trained on one speaker: 34/44
- ▶ Viterbi algorithm accuracy
trained on all speakers: 44/44
trained on one speaker: 34/44 (same mistakes as ML)
- ▶ Computational complexity
Viterbi scoring: $\mathcal{O}(n^2 T)$
Forward scoring: $\mathcal{O}(n^2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

5.3 Viterbi Approximation



- ▶ $\alpha_n(j) = P(x_0, \dots, x_n, Z_n = s_j \mid \theta)$
- ▶ $\log V_n(j) = \max_{i=0}^{M-1} (\log V_{n-1}(i) + \log a_{ij}) + \log \phi_j(x_n)$
- ▶ $B_n(j) = \operatorname{argmax}_{i=0}^{M-1} (\log V_{n-1}(i) + \log a_{ij})$

5.4 Backward Algorithm



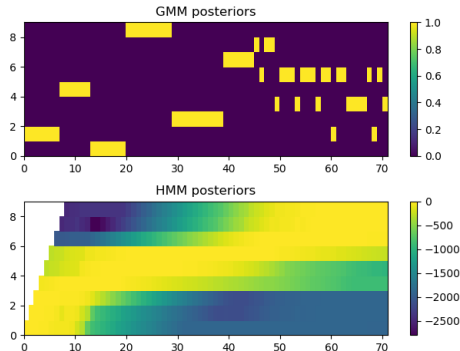
► $\beta_n(i) = P(x_{n+1}, \dots, x_{N-1} \mid z_n = s_i, \theta)$



6.1 State posterior(Gamma) probabilities

- ▶ HMM posteriors: $\gamma_n(i) = P(z_n = s_i \mid X, \theta)$
Summing along state: 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)

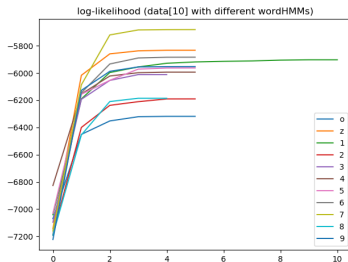
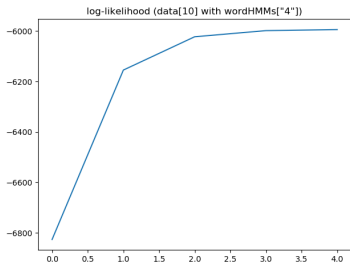
6.1 State posterior(Gamma) probabilities



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



6.2 Retraining the emission probabilities distributions





Thank you for your Attention!