UniS Centre for Vision Speech & Signal Processing
University of Surrey, Guildford GU2 7XH.

HMM tutorial 4

by Dr Philip Jackson

- Discrete & continuous HMMs
  - Gaussian & other distributions
- Gaussian mixtures
  - Revised B-W formulae
- Implementing B-W re-estimation
  - Forward-backward algorithm
  - Accumulation & update
- Summary

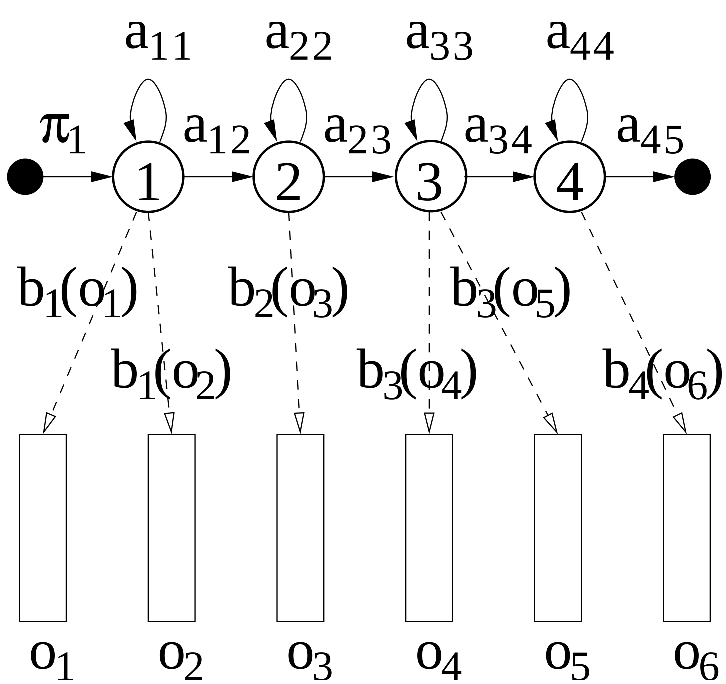
http://www.ee.surrey.ac.uk/Personal/P.Jackson/tutorial/

Discrete HMM, λ

(a) Initial-state probabilities,
$\pi = \{\pi_i\} = \{P(x_1 = i)\}$ for $1 \leq i \leq N$;

(b) State-transition probabilities,
$A = \{a_{ij}\} = \{P(x_t = j | x_{t-1} = i)\}$ for $1 \leq i, j \leq N$;

(c) Discrete output probabilities,
$B = \{b_i(k)\} = \{P(o_t = k | x_t = i)\}$ for $1 \leq i \leq N$ and $1 \leq k \leq K$.



producing *discrete* observations with a state sequence $X = \{1, 1, 2, 3, 3, 4\}$.

Discrete output pdfs

Discretised observations

Continuous HMM, λ

(a) Initial-state probabilities,
$\pi = \{\pi_i\} = \{P(x_1 = i)\}$ for $1 \leq i \leq N$;

(b) State-transition probabilities,
$A = \{a_{ij}\} = \{P(x_t = j | x_{t-1} = i)\}$ for $1 \leq i, j \leq N$;

(c) Continuous output probabilities,
$B = \{b_i(o_t)\} = \{P(o_t | x_t = i)\}$ for $1 \leq i \leq N$,

where the output probability for each state,

$$b_i(o_t) = f\left(o_t; \kappa_i\right), \tag{1}$$

is a function of the observations $f(o_t)$ that depends some model parameters κ_i.

Gaussian output pdfs

Univariate Gaussian (scalar observations)

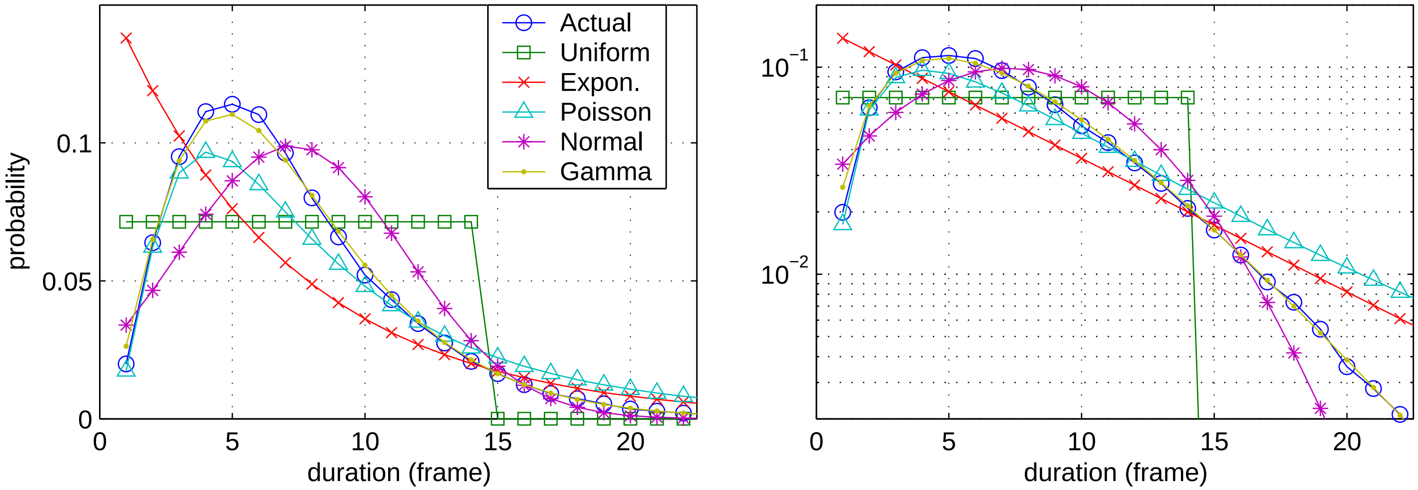
$$b_i(o_t) = \frac{1}{\sqrt{2\pi\Sigma_i}} \exp\left[-\frac{(o_t - \mu_i)^2}{2\Sigma_i}\right].$$

Multivariate Gaussian (vector observations)

$$b_i(\mathbf{o}_t) = \frac{1}{\sqrt{(2\pi)^K |\mathbf{\Sigma}_i|}} \exp\left[-\frac{1}{2}(\mathbf{o}_t - \boldsymbol{\mu}_i)\mathbf{\Sigma}_i^{-1}(\mathbf{o}_t - \boldsymbol{\mu}_i)'\right],$$

where K here refers to the dimensionality of the observation space.

Other distributions



Various functions on (left) linear and (right) logarithmic scales.

- Exponential
- Poisson
- Gamma
- Log-normal
- Gaussian mixture
- Ricean
- Rayleigh
- Beta
- student-t
- Cauchy

Gaussian mixture pdfs

Univariate Gaussian mixture

$$
\begin{aligned}
b_i(o_t) &= \sum_{m=1}^{M} c_{im} \mathcal{N}\left(o_t; \mu_{im}, \Sigma_{im}\right) \\
&= \sum_{m=1}^{M} \frac{c_{im}}{\sqrt{2\pi\Sigma_{im}}} \exp\left[-\frac{(o_t - \mu_{im})^2}{2\Sigma_{im}}\right],
\end{aligned}
\qquad (2)
$$

where M is the number of mixture components in the Gaussian mixture (aka. M-mix), and $\sum_{m=1}^{M} c_{im} = 1$.

Multivariate Gaussian mixture

$$
b_i(\mathbf{o}_t) = \sum_{m=1}^{M} c_{im} \mathcal{N}\left(\mathbf{o}_t; \boldsymbol{\mu}_{im}, \boldsymbol{\Sigma}_{im}\right). \qquad (3)
$$

Revised Baum-Welch formulae

Likelihood of mixture occupation

$$\gamma_t(j,m) = \frac{\alpha_t^-(j)\; c_{jm}\mathcal{N}\left(\mathbf{o}_t; \boldsymbol{\mu}_{jm}, \boldsymbol{\Sigma}_{jm}\right)\; \beta_t(j)}{P(\mathcal{O}|\lambda)} \tag{4}$$

where

$$\alpha_t^-(j) \;=\; \left[\sum_{i=1}^{N} \alpha_t(i) a_{ij}\right]$$

and, as before,

$$\begin{aligned} \gamma_t(j) \;&=\; \frac{\alpha_t(j)\, \beta_t(j)}{P(\mathcal{O}|\lambda)} \\ &=\; \sum_{m=1}^{M} \gamma_t(j,m). \end{aligned} \tag{5}$$

Re-estimation of mixture parameters

For the means

$$\hat{\boldsymbol{\mu}}_{jm} = \frac{\sum_{t=1}^{T} \gamma_t(j,m)\mathbf{o}_t}{\sum_{t=1}^{T} \gamma_t(j,m)}, \tag{6}$$

for the variances

$$\hat{\boldsymbol{\Sigma}}_{jm} = \frac{\sum_{t=1}^{T} \gamma_t(j,m)(\mathbf{o}_t - \boldsymbol{\mu}_{jm})(\mathbf{o}_t - \boldsymbol{\mu}_{jm})'}{\sum_{t=1}^{T} \gamma_t(j,m)}, \tag{7}$$

and for the mixture weights

$$\hat{c}_{jm} = \frac{\sum_{t=1}^{T} \gamma_t(j,m)}{\sum_{t=1}^{T} \gamma_t(j)}. \tag{8}$$

Implementating B-W re-estimation

Overview

- **Forward**: compute likelihoods $\alpha_t(i)$
- **Backward**: compute likelihoods $\beta_t(i)$
- **Parallel**:
  - compute $\gamma_t(i)$ (occupation)
  - compute $\xi_t(i,j)$ (transition)
  - accumulate $\underline{\pi}_i$, $\underline{a}_{ij}$ and $\bar{a}_i$
  - accumulate $\underline{\mu}_i$, $\underline{\Sigma}_i$ and $\bar{b}_i$
- *Repeat* for all files in the training set, $r \in \{1, 2, \ldots, R\}$.

Forward procedure (Problem 1)

1. Initially,
   $\alpha_1(i) = \pi_i \, b_i(o_1)$, for $1 \le i \le N$;

2. For $t = 2, 3, \ldots, T$,
   $\alpha_t(j) = \left[\sum_{i=1}^{N} \alpha_{t-1}(i) \, a_{ij}\right] b_j(o_t)$, for $1 \le j \le N$;

3. Finally,
   $P(\mathcal{O}|\lambda) = \sum_{i=1}^{N} \alpha_T(i)$.

Backward procedure (Problem 1)

1. Initially,
   $\beta_T(i) = 1$, for $1 \le i \le N$;

2. For $t = T-1, T-2, \ldots, 1$,
   $\beta_t(i) = \sum_{j=1}^{N} a_{ij} \, b_j(o_{t+1}) \, \beta_{t+1}(j)$, for $1 \le i \le N$;

3. Finally,
   $P(\mathcal{O}|\lambda) = \sum_{i=1}^{N} \pi_i \, b_i(o_1) \, \beta_1(i)$.

Parallel procedure

Occupation likelihoods:

$$\gamma_t(j) = \frac{\alpha_t(j)\,\beta_t(j)}{P(\mathcal{O}|\lambda)}$$

Transition likelihoods:

$$\xi_t(i,j) = \frac{\alpha_{t-1}(i)\;\,a_{ij}\,b_j(o_t)\;\,\beta_t(j)}{P(\mathcal{O}|\lambda)}$$

Transition accumulators:

$$\begin{aligned}
\underline{\pi}_i &= \dot{\underline{\pi}}_i + \gamma_1(i) \\
\underline{a}_{ij} &= \dot{\underline{a}}_{ij} + \sum_{t=2}^{T} \xi_t(i,j) \\
\bar{a}_i &= \dot{\bar{a}}_i + \sum_{t=2}^{T} \gamma_t(i)
\end{aligned}$$

where $\dot{x}$ denotes to previous value of accumulator x.

Parallel (continued)

Output accumulators (eqs. 11 & 12, from tut. 3):

$$\underline{\mu}_i = \dot{\underline{\mu}}_i + \sum_{t=1}^{T} \gamma_t(i) o_t$$

$$\underline{\Sigma}_i = \dot{\underline{\Sigma}}_i + \sum_{t=1}^{T} \gamma_t(i)(o_t - \mu_i)(o_t - \mu_i)'$$

$$\bar{b}_i = \dot{\bar{b}}_i + \sum_{t=1}^{T} \gamma_t(i)$$

Repeat

For all files in the training set:

1. recompute the forward and backward likelihoods;
2. recompute the occupation and transition likelihoods;
3. increment the accumulators.

Update (Problem 3)

Finally, we update the models:

$$\widehat{\pi}_i = \frac{1}{R} \sum_{r=1}^{R} \underline{\pi_i}$$

$$\widehat{a}_{ij} = \frac{\sum_{r=1}^{R} \underline{a_{ij}}}{\sum_{r=1}^{R} \bar{a}_i}$$

$$\widehat{b}_i \begin{cases} \widehat{\mu}_i = \dfrac{\sum_{r=1}^{R} \underline{\mu_i}}{\sum_{r=1}^{R} \bar{b}_i} \\ \widehat{\Sigma}_i = \dfrac{\sum_{r=1}^{R} \underline{\Sigma_i}}{\sum_{r=1}^{R} \bar{b}_i} \end{cases}$$

Today's summary

- Discrete & continuous HMMs
  - Discrete output pdfs, $b_i(k)$
  - Continuous output pdfs, $b_i(o_t)$
  - Gaussian & other distributions
- Gaussian mixtures, $\sum_m c_{im} \mathcal{N}(\boldsymbol{\mu}_{im}, \boldsymbol{\Sigma}_{im})$
  - Revised B-W formulae
- Implementing B-W re-estimation
  - Forward-backward algorithm
  - Accumulation & update

Next time

- A case study and some illustrative examples
- More implementation details
- Extensions of the HMM framework

Homework

For your data:

- look at the distributions for each of the states in your model and choose a number of mixture components;
- re-train your model, checking for increased likelihood with new model.

As before, use Viterbi to test whether the updated parameters change the state alignment.