Machine learning: lecture 18

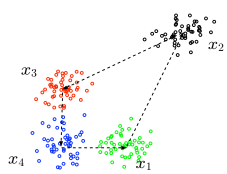
Tommi S. Jaakkola
MIT CSAIL
tommi@csail.mit.edu

Topics

- Observations and Markov properties
- Hidden Markov models (HMMs)
  - examples, problems
  - forward-backward probabilities

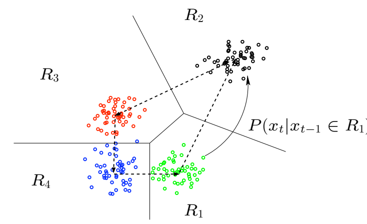
Markov properties: example

- Data generation: successive states x_1, x_2, x_3, and x_4 come from different clusters (colored) in a counter-clockwise manner.



Markov properties: example

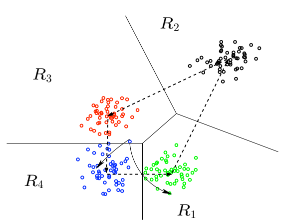
- Data generation: successive states x_1, x_2, x_3, and x_4 come from different clusters (colored) in a counter-clockwise manner.



  - we could try to define state transitions by partitioning the space into regions that contain the clusters

Markov properties: example

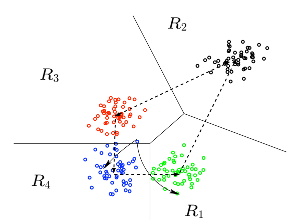
- Data generation: successive states x_1, x_2, x_3, and x_4 come from different clusters (colored) in a counter-clockwise manner.



  - we could try to define state transitions by partitioning the space into regions that contain the clusters
  - problem: overlapping clusters lead to ambiguity

Markov properties: example

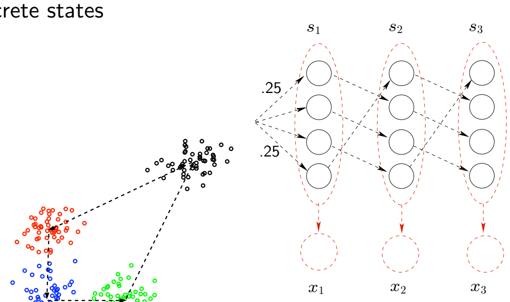
- Data generation: successive states x_1, x_2, x_3, and x_4 come from different clusters (colored) in a counter-clockwise manner.



  - we could try to define state transitions by partitioning the space into regions that contain the clusters
  - problem: overlapping clusters lead to ambiguity
  - we can resolve such ambiguities by looking at previous states (regions visited by the states)

Markov properties: example

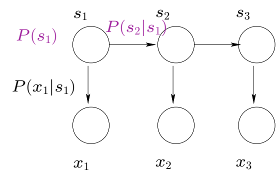
- We can solve the problem more easily by modeling the observations as samples from distributions associated with discrete states

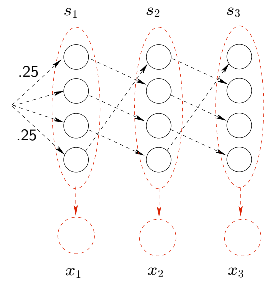


$P(x_1|s_1 = i) = N(x_1; \mu_i, \Sigma_i)$

Hidden Markov Models (HMMs)

- Hidden Markov models are Markov models with state dependent observations

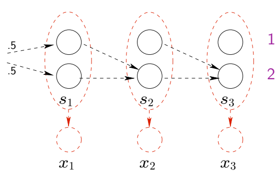




$P(x_1|s_1 = i) = N(x_1; \mu_i, \Sigma_i)$

HMM examples

- Two states 1 and 2; observations are tosses of unbiased coins

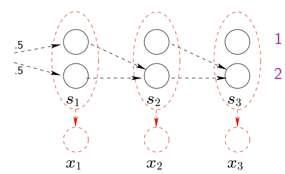


$$P_x(x = \text{heads}|s = 1) = 0.5, \quad P_x(x = \text{tails}|s = 1) = 0.5$$
$$P_x(x = \text{heads}|s = 2) = 0.5, \quad P_x(x = \text{tails}|s = 2) = 0.5$$

- This model is *unidentifiable* from $x-$observations alone

HMM examples: biased outputs

- Two states 1 and 2; outputs are tosses of *biased* coins



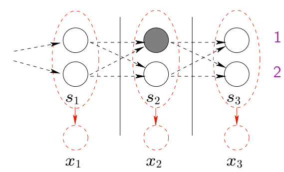
$$P_x(x = \text{heads}|s = 1) = 0.25, \quad P_x(x = \text{tails}|s = 1) = 0.75$$
$$P_x(x = \text{heads}|s = 2) = 0.75, \quad P_x(x = \text{tails}|s = 2) = 0.25$$

- What type of output sequences do we get from this HMM model?

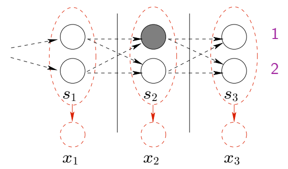
HMM problems

- There are several problems we have to solve
  1. How do we evaluate the probability of an observation sequence $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$?
     – forward-backward algorithm
  2. How do we adapt the parameters of the HMM to better account for the observations?
     – the EM-algorithm
- How do we uncover the most likely hidden state sequence corresponding to the observations?
  – dynamic programming (Viterbi algorithm)

Forward-backward probabilities



Forward-backward probabilities



- Forward (predictive) probabilities $\alpha_t(i)$:

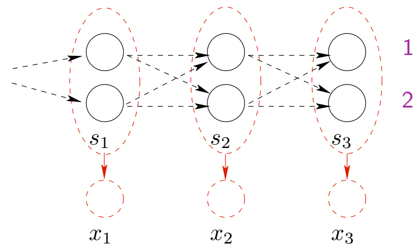
$$\begin{aligned}\alpha_t(i) &= P(\mathbf{x}_1, \ldots, \mathbf{x}_t, s_t = i) \\ \frac{\alpha_t(i)}{\sum_j \alpha_t(j)} &= P(s_t = i | \mathbf{x}_1, \ldots, \mathbf{x}_t)\end{aligned}$$

- Backward (diagnostic) propabilities $\beta_t(i)$:

$$\beta_t(i) = P(\mathbf{x}_{t+1}, \ldots, \mathbf{x}_n | s_t = i)$$

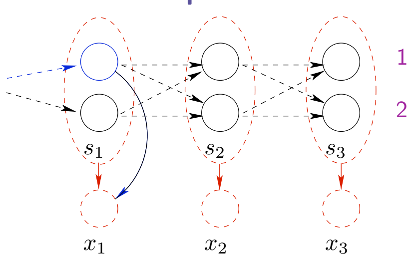
(evidence about the current state from future observations)

Forward probabilities



$\alpha_1(1) = P(x_1, s_1 = 1)$
$\alpha_1(2) = P(x_1, s_1 = 2)$

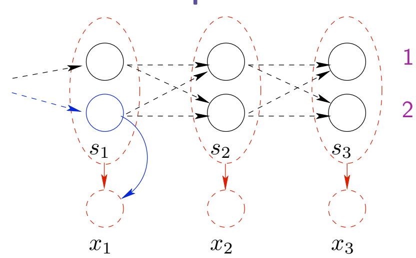
Forward probabilities



$\alpha_1(1) = P(x_1, s_1 = 1)$
$\alpha_1(2) = P(x_1, s_1 = 2)$

$$\alpha_1(1) = P(1) P_x(\mathbf{x}_1 | 1)$$

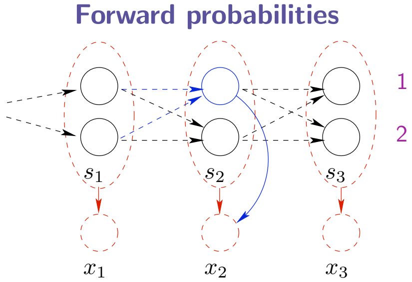
Forward probabilities



$\alpha_1(1) = P(x_1, s_1 = 1)$
$\alpha_1(2) = P(x_1, s_1 = 2)$

$$\begin{aligned}\alpha_1(1) &= P(1) P_x(\mathbf{x}_1 | 1) \\ \alpha_1(2) &= P(2) P_x(\mathbf{x}_1 | 2)\end{aligned}$$

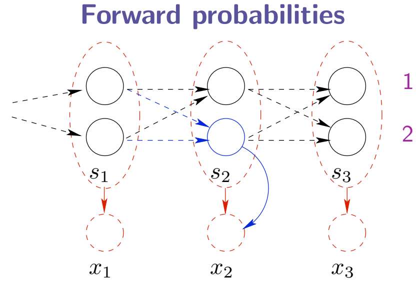
Forward probabilities



$\alpha_2(1) = P(x_1, x_2, s_2 = 1)$
$\alpha_2(2) = P(x_1, x_2, s_2 = 2)$

$$\alpha_2(1) = \big[\alpha_1(1) P_1(1|1) + \alpha_1(2) P_1(1|2)\big] P_x(\mathbf{x}_2 | 1)$$

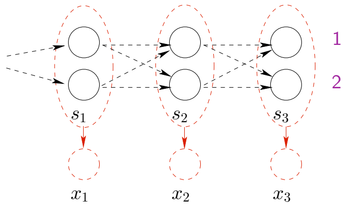
Forward probabilities



$\alpha_2(1) = P(x_1, x_2, s_2 = 1)$
$\alpha_2(2) = P(x_1, x_2, s_2 = 2)$

$$\begin{aligned}\alpha_2(1) &= \big[\alpha_1(1) P_1(1|1) + \alpha_1(2) P_1(1|2)\big] P_x(\mathbf{x}_2 | 1) \\ \alpha_2(2) &= \big[\alpha_1(1) P_1(2|1) + \alpha_1(2) P_1(2|2)\big] P_x(\mathbf{x}_2 | 2)\end{aligned}$$

Forward probabilities

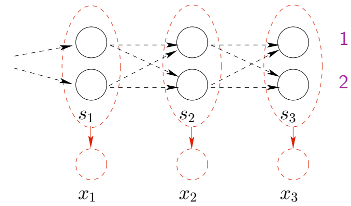


- We get the following recursive equation for calculating the forward probabilities $\alpha_t(i) = P(\mathbf{x}_1, \ldots, \mathbf{x}_t, s_t = i)$:

$$\alpha_1(i) = P(i) P_x(\mathbf{x}_1|i)$$

$$\alpha_t(i) = \left[\sum_j \alpha_{t-1}(j) P_1(i|j) \right] P_x(\mathbf{x}_t|i)$$

Backward probabilities



- We can proceed analogously to derive a recursive equation for the backward probabilities $\beta_t(i) = P(\mathbf{x}_{t+1}, \ldots, \mathbf{x}_n | s_t = i)$:

$$\beta_n(i) = 1$$

$$\beta_t(i) = \left[\sum_j P_1(j|i) P_x(\mathbf{x}_{t+1}|j) \beta_{t+1}(j) \right]$$

Uses of forward/backward probabilities

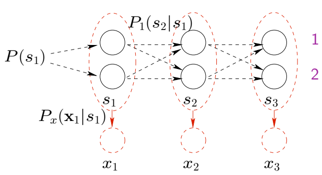
- The complementary forward/backward probabilities

$$\alpha_t(i) = P(\mathbf{x}_1, \ldots, \mathbf{x}_t, s_t = i)$$

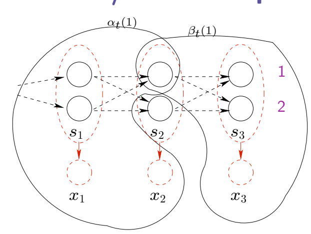
$$\beta_t(i) = P(\mathbf{x}_{t+1}, \ldots, \mathbf{x}_n | s_t = i)$$

permit us to evaluate various probabilities:

1. $P(\mathbf{x}_1, \ldots, \mathbf{x}_n)$
2. $\gamma_t(i) = P(s_t = i | \mathbf{x}_1, \ldots, \mathbf{x}_n)$
3. $\xi_t(i, j) = P(s_t = i, s_{t+1} = j | \mathbf{x}_1, \ldots, \mathbf{x}_n)$



Uses of forward/backward probabilities



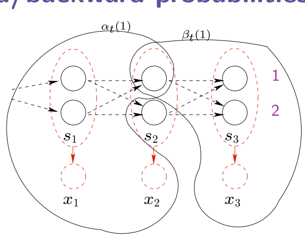
Probability of the observation sequence:

$$P(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \sum_i P(\mathbf{x}_1, \ldots, \mathbf{x}_n, s_t = i)$$

$$= \sum_i P(\mathbf{x}_1, \ldots, \mathbf{x}_t, s_t = i) P(\mathbf{x}_{t+1}, \ldots, \mathbf{x}_n | s_t = i)$$

$$= \sum_i \alpha_t(i) \beta_t(i)$$

Forward/backward probabilities cont'd



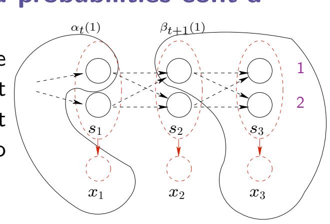
- We can evaluate the posterior probability that the HMM was in a particular state i at time t

$$P(s_t = i | \mathbf{x}_1, \ldots, \mathbf{x}_n) = \frac{P(\mathbf{x}_1, \ldots, \mathbf{x}_n, s_t = i)}{P(\mathbf{x}_1, \ldots, \mathbf{x}_n)}$$

$$= \frac{\alpha_t(i) \beta_t(i)}{\sum_j \alpha_t(j) \beta_t(j)} \stackrel{def}{=} \gamma_t(i)$$

Forward/backward probabilities cont'd

- We can also compute the posterior probability that the system was in state i at time t AND transitioned to state j at time $t+1$:



$$P(s_t = i, s_{t+1} = j | \mathbf{x}_1, \ldots, \mathbf{x}_n)$$

$$= \frac{\alpha_t(i) \overbrace{P_1(s_{t+1} = j | s_t = i) P_x(\mathbf{x}_{t+1} | s_{t+1} = j)}^{\text{fixed } i \rightarrow j \text{ transition, one observation}} \beta_{t+1}(j)}{\sum_j \alpha_t(j) \beta_t(j)}$$

$$\stackrel{def}{=} \xi_t(i, j),$$