Lecture 12 More Examples

Thursday, September 29, 2016 10:43 PM

- Hidden Markov Model (HMM)
  - model of time series data

    Markov property: what happens next only depends on what happens now.

    Often a bit restrictive if we don't observe everything

  - HMM: hidden state $h_t \in \{1, 2, \cdots, r\}$

    $t = 1, 2, \cdots$ denotes time

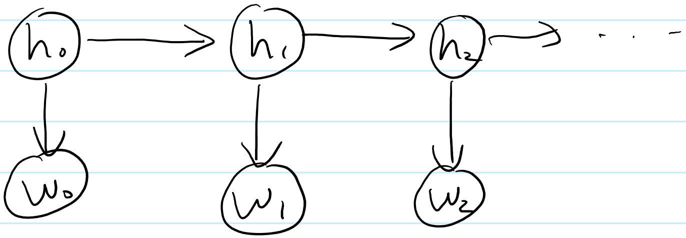
    "transition" matrix P

    $$\Pr[h_{t+1} = i \mid h_t = j] = P_{ij}$$

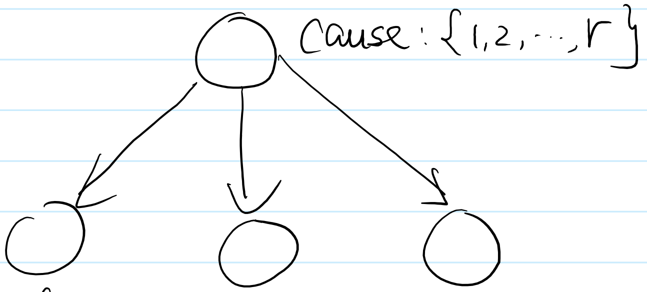
    observations $w_t \in \{1, 2, \cdots, n\}$

    observation matrix Q

    $$\Pr[w_t = i \mid h_t = j] = Q_{i,j}$$



- multi-view interpretation
  - rank 1 tensor $\Longleftrightarrow$ 3 independent variables
  - rank r tensor $\Longleftrightarrow$ 3 variables that are independent conditioned on 1 of r "causes"



try to find this structure in the model!

- for HMM: w_0, w_1, w_2 are independent conditioned on h_1

$$\Pr[w_0, w_1, w_2] = \sum_{i=1}^{r} \Pr[w_0|h_i]\Pr[w_1|h_i]\Pr[w_2|h_i]\Pr[h_i]$$

low rank tensor!

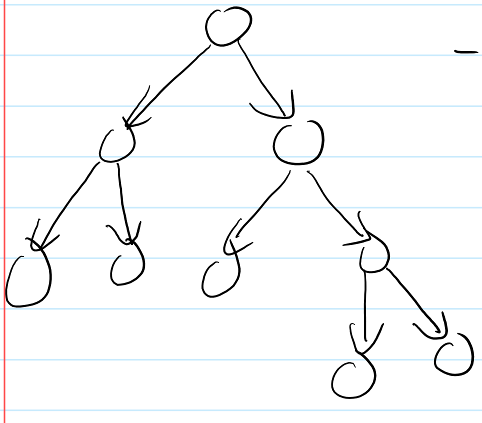
$$\Pr[w_1|h_i] = Q \quad \Pr[w_2|h_i] = QP$$

using these information, can recover Q, P if both have rank r

- Extension: Phylogenic tree reconstruction
  - model for evolution of species
    - for each species, n locations in DNA sequence
    - assumption: these locations are "useless" and independent (probably not true)

      they evolve according to a random transition matrix

      ($\Pr[A \to G]$ ... 4×4)



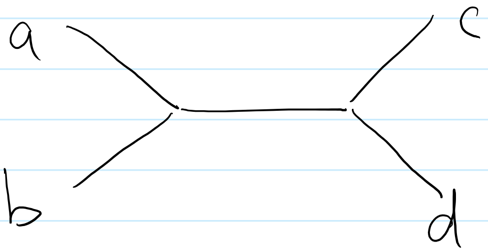
- phylogenetic tree
  - nodes = species
  - branches: one species separates into two species
  - only leaves are observed

    internal nodes correspond to extinct species.

- Goal: ① Find the tree structure

  ② Learn the transition matrix

– ① : use "Quartet test": measure "distance" between nodes



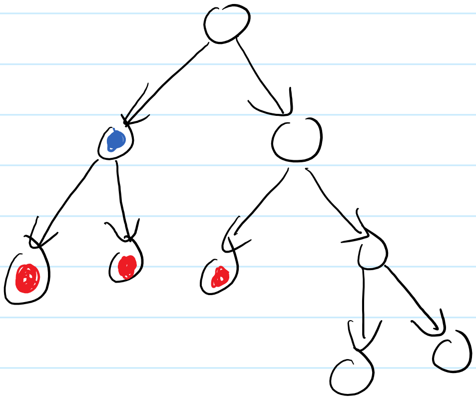
in this case

$$d(a,b) + d(c,d) < d(a,c) + d(b,d)$$

$$d(a,b) + d(c,d) < d(a,d) + d(b,c)$$

For every 4 leaves, can find out which pair "met" first in the tree. Possible to construct the tree based on these information.

– ②



Notice: the three red leaves are independent conditioned on blue node.

Can use tensor decomposition to learn the transition from blue to any red leaf.