Training HMMs

Goal: Given: • Long sequence of observations $X = x_1 \dots x_L$
• Structure of an HMM: Set of states S
Set of possible trans. T

Find: • Emission probabilities : E
• Transition prob. : T
• Initial state prob : I

such that E, T, I best represent the observed seq. X

Maximum Likelihood estimation ← $\Pr[X = x_1 \dots x_L \mid E, T, I]$ is maximized

$$\Pr[A] = \sum_b \Pr[A, B = b]$$

$$\Pr[X = x_1 \dots x_L \mid E, T, I] = \sum_{\text{path } P = p_1 \dots p_L} \Pr[X, P \mid E, T, I]$$

We know how to calculate

Can be calculated in $O(L \cdot n^2)$ using Forward algorithm

Simplified version:

Given: $X = x_1 \dots x_L$
$P = P_1 \dots P_L$ ← annotation of X

Find: E, T, I s.t. $Pr[X, P \mid E, T, I]$ is max

Let $N_t(s, s')$ = # of times transition happened btw s and s'
= # of positions i s.t. $P_i = s \wedge P_{i+1} = s'$

Then, choose $T(s, s') = \dfrac{N_t(s, s')}{\sum_{x \in S} N_t(s, x)}$ → total # of times in state s

Example: X: A C A ' I A C I A T G A C T G
P: G G G I I I I I I G G G G I I

$T(G, I) = \frac{2}{7}$ $\quad T(G, G) = \frac{5}{7}$

For emissions: $N_e(s, a)$ = # times a was emitted from s
= # positions i s.t. $P_i = s \wedge X_i = a$

Choose $E(s, a) = \dfrac{N_e(s, a)}{\sum_b N_e(s, b)}$

Example: $E(G, A) = \frac{4}{7}$

Back to problem where P is not given

① Viterbi training

Ⓐ Choose "reasonable" E, T, I (either from prior knowledge or randomly)

Repeat until convergence

Ⓑ Use Viterbi algo to find best path for $X = x_1 \ldots x_k$ assuming E, T, I ⇒ Produces path P

Ⓒ Re-estimate E, T, I from P, X as done previously

Problem: Algo. frequently gets stuck in local optima

② Baum-Welsch algorithm

Bases re-estimation of E, T, I on all paths (weighted by their probability)

⇒ Reduces the prob of getting stuck in local optima