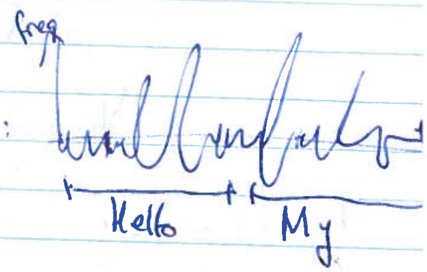
Hidden Markov Models (HMM)

Goal: "Decode" a sequence of observations

Speech recognition : Observation: 

Biological sequence
Gene annotation : Observation: ACTACTAGGTAC
Genes

Protein: VLIYVRRCY.

α-helix β-str

Toy example: You visit a city with different neighborhoods

set of states = $S = \{F, E, C\} = S = \{S_1, S_2, \ldots S_n\}$

French English Chinese

You walk randomly in city.

Every minute, someone greets you.

set of symbols = $\Sigma = \{b, h, n, a\} = \{\sigma_1, \sigma_2, \ldots \sigma_k$

bonjour hello chinese namaste

You record seq. of greetings: $X = x_1, x_2, \ldots x_L$

where $x_i \in \Sigma$ $X = bbaahhhahb\ldots$

Problem: Given: $X = x_1 \ldots x_n$

Find: Path $P = p_1 \ldots p_n$ that is most likely given X

We need to know:

① Emission Probabilities: $Pr_e[x_i = \alpha \mid p_i = \beta]$ ($\alpha \in \Sigma$, $\beta \in S$)

symbols

States	b	h	n	a
F	0.9	0.05	0.05	0
E	0.1	0.5	0.1	0.3
C	0.3	0.3	0.3	0.1

→ sum to 1

② Transition probabilities

$Pr_t[P_{i+1} = \gamma \mid P_i = \beta]$

$\in S$ $\in S$

Destination

Source	E	C	F
E	0.9	0	0.1
C	0.1	0.4	0.5
F	0.3	0.1	0.6

Path: E EEEEEEE FFFFEE FFFCCC

③ Initial State probability

$Pr[P_1 = \gamma]$

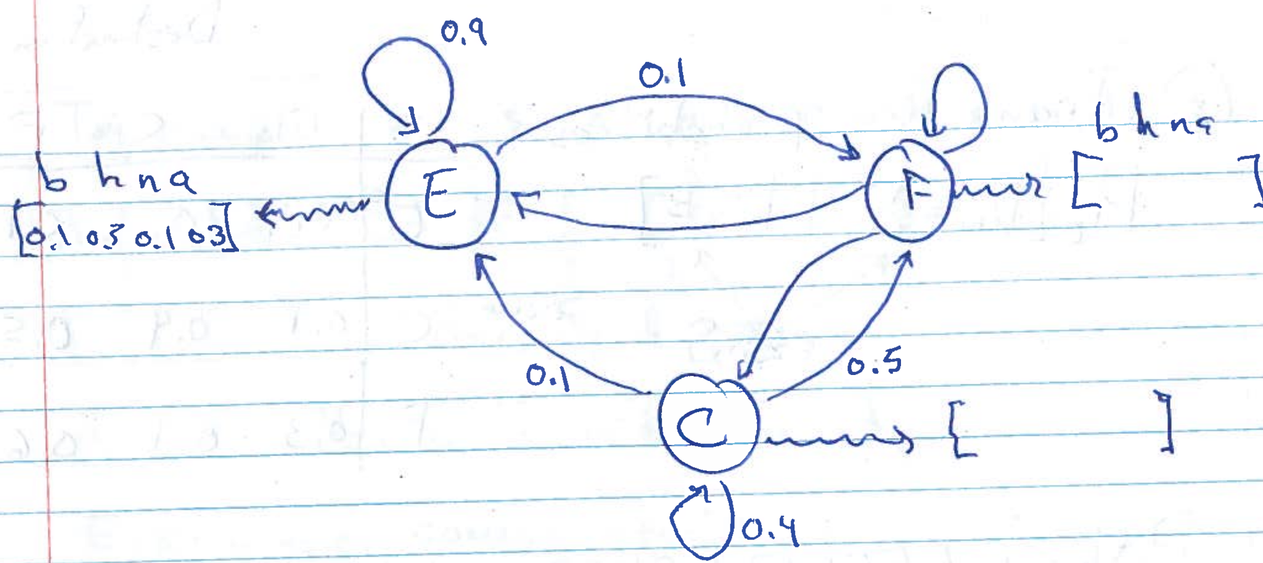
$\in S$

E	C	F
0.5	0	0.5

Assumptions: - Markovian assumption (1st order)

Prob of going to state γ only depends on where you are now, not on how you got there

- Observations are independent from each other (given the states)



HMM as generative model: Generate a random obs sequence

① Pick randomly P_1 from the initial state prob. dist

Repeat: 2.1 Emit an observation from current state

2.2 Transition to new state acc. to transition prob. matrix

Questions (Assume we know S, Σ, emission prob, transition prob, initial prob)

① Maximum likelihood path:
Given: $X = x_1 \ldots x_L$
Find: Path $P = P_1 \ldots P_L$ that is most likely to have generated X, ie $Pr[P_1 \ldots P_L \mid x_1 \ldots x_L]$ is maximized

Answer: Viterbi algo.

② Posterior decoding
Given: $X = x_1 \ldots x_L$, time i, state β
Calculate: $Pr[P_i = \beta \mid X]$

Answer: Forward-Backward algo

③ Estimation Problem: Assume we know S, Σ
but not emission prob
not transition prob
not initial

Given: $X = x_1 \dots x_L$

Find: Emission prob, Transition prob, Initial prob $\Big\}$ such that $\Pr[X \mid E, T, I]$ is max

Answer: Baum-Welch algorithm

Viterbi Algorithm

Given: $X = x_1 \dots x_L$ (seq. of L observations)

S, Σ, E, T, I = HMM

states, alphabet, emission prob matrix, Transition prob matrix, initial state prob. vector

Find: $P = P_1 \dots P_L$, where $P_i \in S$

such that $\Pr[P_1 \dots P_L \mid X = x_1 \dots x_n]$ is maximized

Question 1: How to calculate $Pr[P_1 \dots P_L \mid X = x_1 \dots x_L]$? for a given path $P = P_1 \dots P_L$

$$Pr[P_1 \dots P_L \mid X = x_1 \dots x_L] = \frac{Pr[P_1 \dots P_L \wedge X = x_1 \dots x_L]}{Pr[X = x_1 \dots x_L]}$$

Independent of Path P

$$Pr[P_1 \dots P_L \wedge X = x_1 \dots x_L] = Pr[X = x_1 \dots x_L \mid P_1 \dots P_L] \cdot Pr[P_1 \dots P_L]$$

$$= \left(\prod_{i=1}^{L} Pr_e[x_i \mid P_i] \right) \cdot Pr_I(P_1) \cdot \prod_{i=1}^{L-1} Pr_t[P_{i+1} \mid P_i]$$

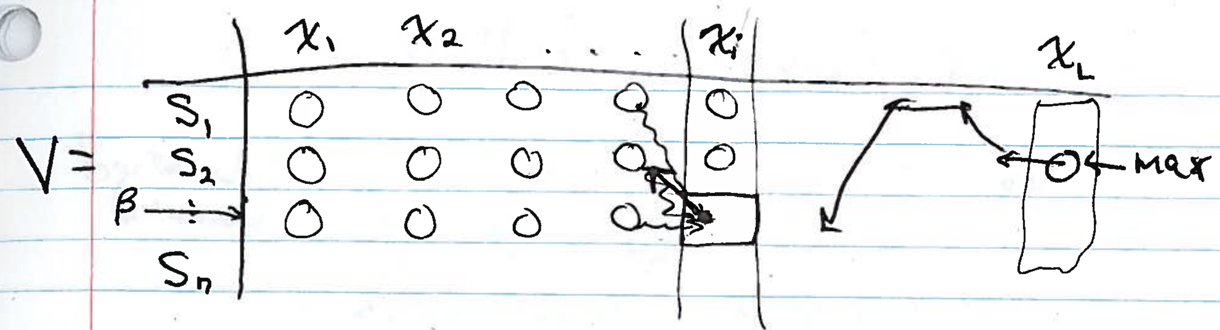
Question 2: How to find P s.t. $Pr[P_1 \dots P_L \mid x_1 \dots x_L]$ is max?

$\Longrightarrow$ Viterbi algorithm

Define $V(\beta, i)$ = Prob. of the most likely path of length i, given observation $x_1 \dots x_i$, assuming path ends in state β

$\beta \in S$, $i \in \{1 \dots L\}$

$$= \max_{\substack{P_1 \dots P_i \\ \text{where } P_i = \beta}} \{ Pr[P_1 \dots P_i, x_1 \dots x_i] \}$$



$$V(\beta, i) = \max_{\gamma \in S} \{ V(\gamma, i-1) \cdot Pr_t[\beta | \gamma] \} \cdot Pr_e[x_i | \beta]$$

Fill V table column by column, left-to-right

$V(\beta, 1) = Pr_I(\beta) \cdot Pr_e[x_1 | \beta]$ (Initialization)

$$\max_{P_1 \ldots P_L} \{ Pr[P_1 \ldots P_L, x_1 \ldots x_L] \} = \max_{\beta \in S} \{ V(\beta, L) \}$$

To recover path P that maximized $Pr[P_1 \ldots P_L, x_1 \ldots x_L]$

- Trace back dyn. prog algo from $\max\{V(\beta, L)\}$

<u>Complexity</u>

Time = $O(n^2 \cdot L)$

Space = $O(n \cdot L)$