Hidden Markov Models

A hidden markov model consists of a set of internal states and a set of observable tokens. A run of a hidden Markov model generates a hidden state sequence $s_1, \ldots, s_T$ and a sequence of observable tokens $a_1, \ldots, a_T$.

$$\pi(s) \;=\; P(s_1 = s),\; \sum_s \pi(s) = 1$$

$$T(w|s) \;=\; P(s_{t+1} = w | s_t = s),\; \sum_w T(w|s) = 1$$

$$O(a|s) \;=\; P(a_t = a | s_t = s),\; \sum_a O(a|s) = 1$$

$$P(s_1, \ldots, s_T, a_1, \ldots, a_T) \;=\; \pi(s_1)\left(\Pi_{t=1}^{T} O(a_t|s_t)\right)\left(\Pi_{t=1}^{T-1} T(s_{t+1}|s_t)\right)$$

Some Applications of HMMs

- Speech Recognition. The hidden states are word positions and the observable tokens are accoustic feature vectors.

- Part of speech tagging. The hidden states are the parts of speech (noun, verb, adjective, and so on).

- DNA sequence analysis. The hidden states might be protein secondary structure or a position in a homologous sequence.

Viterbi Algorithm

$$\mathrm{Viterbi}[\mathrm{s_t}, \mathrm{t}] = \max_{s_1,\ldots,s_{t-1}} P(s_1, \ldots, s_t, a_1, \ldots, a_{t-1})$$

$$\mathrm{Viterbi}[\mathrm{s_1}, 1] = \pi(s_1)$$

$$\mathrm{Viterbi}[\mathrm{s_{t+1}}, \mathrm{t}+1] = \max_{s_t} \ \mathrm{Viterbi}[\mathrm{s_t}, \mathrm{t}]\mathrm{O}(\mathrm{a_t}|\mathrm{s_t})\mathrm{T}(\mathrm{s_{t+1}}|\mathrm{s_t})$$

$$s_T^* = \operatorname*{argmax}_s \ \mathrm{Viterbi}(\mathrm{s_T}, \mathrm{T})\mathrm{O}(\mathrm{a_T}|\mathrm{s_T})$$

The best predicessor s_t can be recorded for each possible value of s_{t+1} and the best path can be constructed by working backward from s_T^* through best predicessors.

Forward-Backward Procedure

$$\text{Forward}[\text{s}_\text{t}, \text{t}] = P(a_1, \ldots, a_{t-1}, s_t)$$
$$\text{Backward}[\text{s}_\text{t}, \text{t}] = P(a_t, \ldots, a_T \mid s_t)$$

$$\text{Forward}[\text{s}_1, 1] = \pi(s)$$
$$\text{Forward}[\text{s}_{\text{t}+1}, \text{t}+1] = \sum_{s_t} \text{forward}[\text{s}_\text{t}, \text{t}]\text{O}(\text{a}_\text{t}|\text{s}_\text{t})\text{T}(\text{s}_{\text{t}+1}|\text{s}_\text{t})$$

$$\text{Backward}[\text{s}_\text{T}, \text{T}] = O(a_T|s_T)$$
$$\text{Backward}[\text{s}_\text{t}, \text{t}] = O(a_t|s_t) \sum_{s_{t+1}} T(s_{t+1}|s_t)Backward[s_{t+1}, t+1]$$

$$P(a_1, \ldots, a_T) = \sum_{s_1} \pi(s_1)\text{Backward}[\text{s}_1, 1]$$
$$P(s_t \mid a_1, \ldots, a_T) = \frac{\text{Forward}(\text{s}_\text{t}, \text{t})\text{Backward}[\text{s}_\text{t}, \text{t}]}{P(a_1, \ldots, a_T)}$$

Problem

Suppose that we have two hidden states A and B and two observable symbols a and b and an HMM defined by the following probabilities.

$$\pi(A) = \pi(B) = 1/2$$

$$T(A|A) = T(B|B) = 1 - \epsilon$$

$$T(B|A) = T(A|B) = \epsilon$$

$$O(a|A) = O(b|B) = 1 - \delta$$

$$O(b|A) = O(a|B) = \delta$$

Now suppose that we observe a sequence of T a's. Let $F(A, t)$ abbreviate Forward(A, t) and similarly for $F(B, t)$ Give the values of $F(A, 1)$ and $F(B, 1)$ and give equation for computing $F(A, t+1)$ and $F(B, t+1)$ as a function of $F(A, t)$ and $F(B, t)$. Similarly let $B(A, t)$ abbreviate Backward(A, t). Give the values of $B(A, T)$ and $B(B, t)$ and give equation for computing $B(A, t)$ and $B(B, t)$ as a function of $F(A, t+1)$ and $F(B, t+1)$.

Extra Credit: Solve for $F(A, t)$, $B(A, t)$ and $P(a_1, \ldots, a_T)$. Graph $P(s_t = A)$ as a function of t for $\epsilon = \delta = 1/4$, and $T = 20$ (you should not calculate the actual numbers for $P(s_t = A)$ if you can see qualitatively what the graph must look like).