Hidden Markov & Maximum Entropy Models

1913 → A.A. Markov : Could we use frequency count from the text to help to compute the probability that the next letter in sequence would be a vowel?

Machine Learning Model
- → Markov Model (MM) / Markov Chain (MC)
- → Hidden Markov Model (HMM)
- → Maximum Entropy Markov Model (MEMM)

HMM & MEMM are both sequence classifier.

Sequence Classifier :

- A model whose job is to assign some label or class to each unit in a sequence.
- Some time this model works with the probabilistic values of the sequence.
- HMM & MEMM are two different probabilistic sequence classifier.

Markov Chain : This is an extension of Finite Automata (FA).

FA → 1) Set of States
2) set of transitions between the states.

Weighted Finite State Automata :-

Each path associated with a probability value such that all path leaving a node must sum to 1.

Markov Chain : A special case of WFSA, in which the input sequence will determines, which states the automaton will go through.

→ This will not represent any un-ambiguous problem.
MC is only useful for representing un-ambiguous problem.

Markov Chain is a probabilistic Graphical Model, consists with following components :

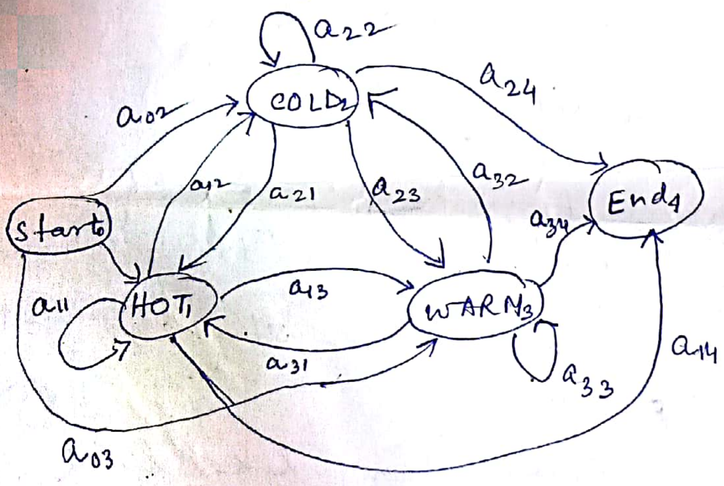
$Q = q_1 q_2 \cdots q_N$ a set of N states

$A = a_{01} a_{02} \cdots a_{n1} \cdots a_{nn}$: a transition probability matrix A. each a_{ij} representing the probability of moving from state i to state j s.t. $\sum_{j=1}^{n} a_{ij} = 1 \; \forall i$

q_0, q_F : a special start state and end (final) state that are not associated with observations.

Markov Assumption: $P(q_i \mid q_1 \cdots q_{i-1}) = P(q_i \mid q_{i-1})$

$$\sum_{j=1}^{n} a_{ij} = 1 \; \forall i$$



A. MC does not rely on a start or end states. Instead representing the distribution over initial states and accepting state.

$\pi = \pi_1, \pi_2, \cdots \pi_n$

↳ an initial probability distribution over states. π_i is the probability that the Markov chain will start with state i. Some states j may have $\pi_j = 0$, meaning that they can not be in initial state. Also $\sum_{i=1}^{n} \pi_i = 1$.

$Q_A = q_x, q_y, \cdots$ a set of legal accepting states. $Q_A \subset Q$

The Hidden Markov Model :-

In many cases, the events (States) we are interested are hidden. We don't observe them directly. E.g., We directly not observed the parts of speech tagged in a text. Rather, we see words, and must infer the tags from the word sequence. Here tags are hidden because they are not observed

Hidden Markov Model (HMM) allows us to talk about both observed events (like words that we see in the input) and hidden events (like parts-of-speech tags) that we think of a causal factors in our probabilistic model.

A HMM contains following components :-

$Q = q_1 q_2 \cdots q_N$ → set of N states

$A = a_{11} \cdots a_{ij} \cdots a_{NN}$ → transition probability matrix A.

$O = o_1 o_2 \cdots o_T$ → a sequence of T observation each one drawn from a vocabulary $V = v_1, v_2, \cdots v_v$.

$B = b_i(o_i)$ → a sequence of observation likelihoods, also called emisson probability each expressing the probability of an observation o_i being generated from a state i.

Scanned by CamScanner