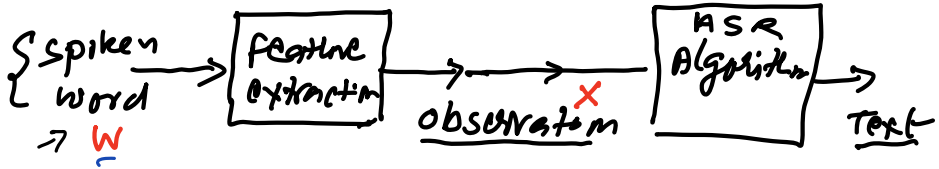
ASR - Automatic speech Recognition.



The algorithm estimates (guesses, or computes based on probability) the spoken word.

What is probability of W given that we have observed the sequence X

$$P(W|X) = \frac{P(X|W)P(W)}{P(X)} \quad \text{— (1)}$$

the word can be $w_1, w_2, \ldots, w_n$. So we will select the word W_i for which the probability in RHS is maximum.

$$W^* = \underset{W}{\text{argmax}}\ P(X|W)P(W) \quad \text{— (2)}$$

W^* is value of W for which (2) is maximum

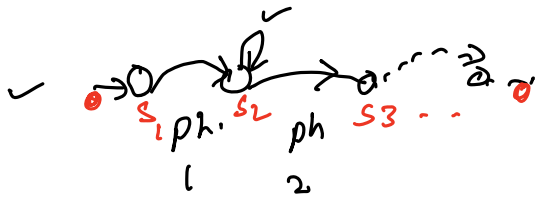
Now we need to compute a) $P(X|W)$: acoustic model

b) $P(W)$: language model

9) Find $P(X|w)$? i.e. probcbity of Feature Vector X given that word is w ?

Say, w = "Look" (No. of phonels = 2)

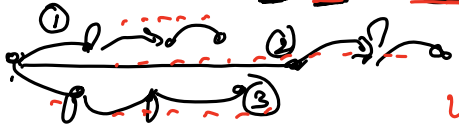
phoneme 1, phoneme 2, there are transitions.



$P(X|w)$ is probability of transitions from state s_1 to s_2, to s_3, ... s_i

$$= \text{Prob. of } s_1 \text{ to } s_2 * \text{Prob. of } s_2 \text{ to } s_3$$

$$P(X|w) = \arg\max_{s_i} \prod P(X|s_i) * P(s_i)$$



] No. of sequences, select the one of max. prob.

In this case there is associated probability with each transition, and the probability of each sequence (of states) is different. Select that sequence of states which has max. probability

Let no. of states = N

and no. of observations = X

$O(N^X)$

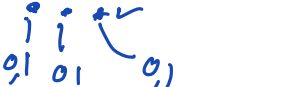
↑ complexity

↑ exponential

∴ we limit the no. of observations to reduce the complexity, e.g. a word spoken may be look, cook, book,

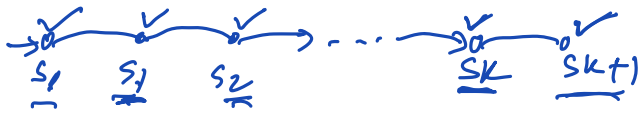
X = {book, look, cook}

If states are $N = \{0, 1\}$ and observations are 3



∴ possible sequences = 2^3 =

let the states are $s_0, s_1, \cdots, s_k, s_{k+1}$



$$P(s_{k+1} \mid s_0, s_1, s_2, \cdots, s_k) = P(s_{k+1} \mid s_k)$$

- it simplifies the computation
- it is used in Markov model of speech recognition