Hidden Markov Model (HMM)

Objective

- Nature of real-world signals
- Modeling real-world signals
- Choosing an appropriate model
  - Deterministic Model
  - Stochastic Model

Stationarity / Nonstationarity

Stationary Process:

Statistical properties do not vary with time

Non-stationary Process:

Statistical properties vary over time

Weather Prediction Example

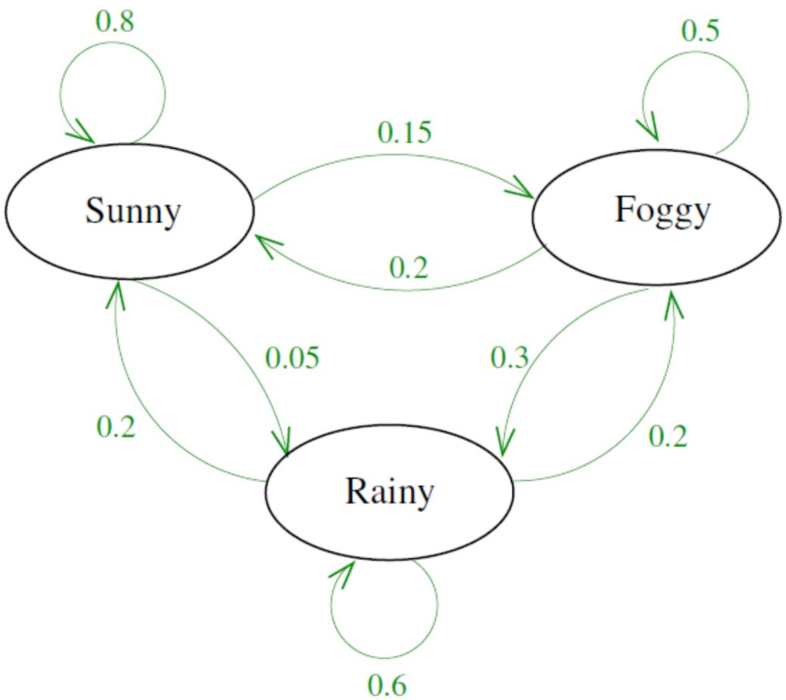
- Predicting todays's weather given the past history
- Given the observed weather condtions for last three days, we are interested in predicting today's.
- P(q4=Rainy/q3=Cloudy, q2=Sunny, q1=Sunny)
- P(qn/qn-1, qn-2,qn-3,….,q1) – Nth order Markov
- We need to collect statistics for pow(3,n-1) histories
- We can simplify the process, by assuming a first order Markov.
- No matter what happens, todays weather depends only on yesterdays weather condition!

First -Order Markov Model

- P(qn/qn-1, qn-2,qn-3,....,q1)=P(qn/qn-1)
- P(q1, q2, ..., qn)=PI(P(qi/qi-1)   i=1,...n
- A discrete (finite) system:
  - N distinct states.
  - Begins (at time t=1) in some initial state(s).
  - At each time step (t=1,2,...) the system moves from current to next state (possibly the same as the current state) according to transition probabilities associated with current state.
- After Andrei Andreyevich Markov (1856 -1922)

Markov Model for Weather Prediction

Today/ Tomorrow	Sunny	Rainy	Cloudy
Sunny	0.8	0.05	0.15
Rainy	0.2	0.6	0.2
Cloudy	0.2	0.3	0.5



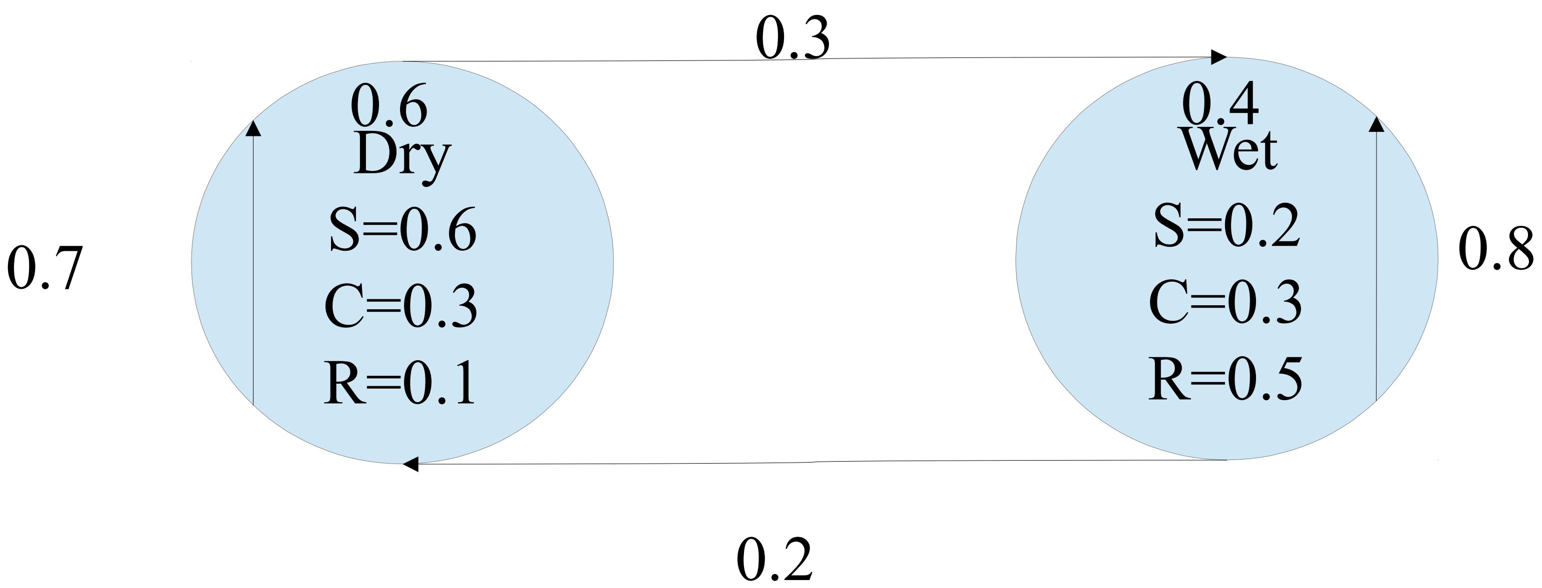
P(q3=S/q2=C,q1=R)=?

P(q2=S, q3=R/q1=S)=P(q3=R/q2=S,q1=S).P(q2=S/q2/S)

=P(q3=R/q2=S).P(q2=S/q1=S)

=0.05 x 0.8 =0.4

Generic Example



P(q2=S, q3=R/Dry)=?

N states and T observations require pow(N,T) computations

HMMs – Question I

- Given an observation sequence $\mathbf{O} = (O_1\ O_2\ O_3\ \ldots\ O_L)$, and a model $M = \{A, B, \pi\}$, how do we efficiently compute $P(O|M)$, the probability that the given model M produces the observation O in a run of length L ?

- This probability can be viewed as a measure of the quality of the model M. Viewed this way, it enables discrimination/selection among alternative models.

Computing Likelihood Using HMM

Option 1) The likelihood is measured using any sequence of states of length T

- This is known as the "Any Path" Method

Option 2) We can choose an HMM by the probability generated using the best possible sequence of states

- We'll refer to this method as the "Best Path" Method

HMM – Question II (Harder)

- Given an observation sequence, $O = (O_1\ O_2\ \dots\ O_T)$, and a model, $M = \{A, B, p\}$, how do we efficiently compute the most probable sequence(s) of states, Q ?
- Namely the sequence of states $Q = (Q_1\ Q_2\ \dots\ Q_T)$, which maximizes $P(O|Q,M)$, the probability that the given model M produces the given observation O when it goes through the specific sequence of states Q .
- Recall that given a model M, a sequence of observations O, and a sequence of states Q, we can efficiently compute $P(O|Q,M)$ (should watch out for numeric underflows)

Most Probable States Sequence (Q. II)

Idea:

- If we know the identity of Q_i, then the most probable sequence on $i+1,\ldots,n$ does not depend on observations before time i
- A white board presentation of Viterbi's algorithm

HMM – Question III (Hardest)

- Given an observation sequence $O = (O_1\ O_2\ \ldots\ O_L)$, and a class of models, each of the form M = {A,B,p}, which specific model "best" explains the observations?
- A solution to question I enables the efficient computation of $P(O|M)$ (the probability that a specific model M produces the observation O).
- Question III can be viewed as a learning problem: We want to use the sequence of observations in order to "train" an HMM and learn the optimal underlying model parameters (transition and output probabilities).

Thank You!