Hidden Markov Models

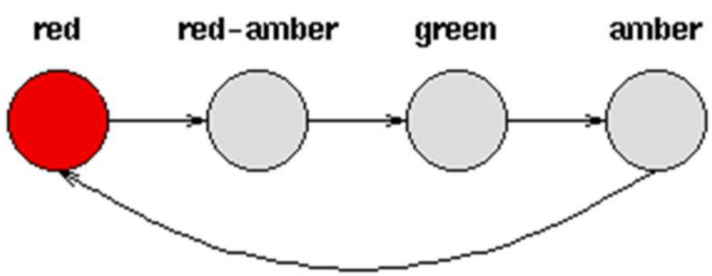
Introduction

- Modeling dependencies in input; no longer independent
- Sequences:
  - Words in a sentence (syntax, semantics of the language)
  - Handwriting: pen movements
  - Speech; phonemes in a word (dictionary)

Patterns

Deterministic Patterns

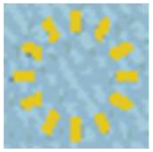
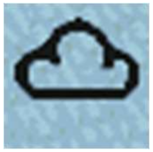
- Traffic Light
- Each state dependent only on the previous state
- System is Deterministic



Patterns

Non-deterministic Patterns

- Weather

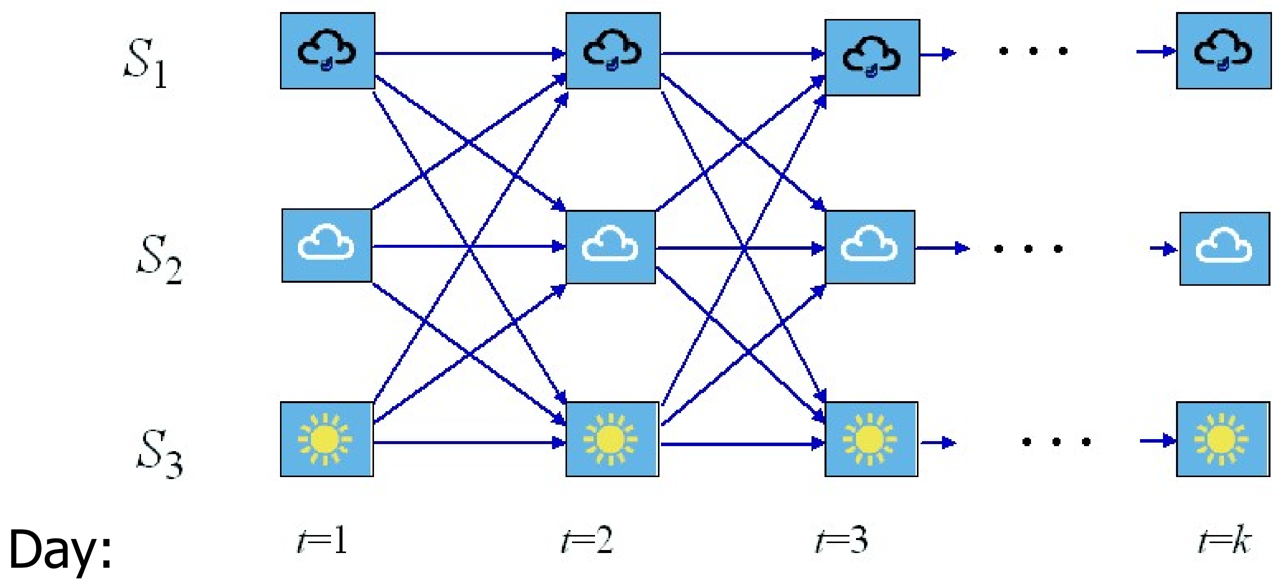
- States do not follow each other deterministically
- Model the system generating these states?
- Markov Assumption – State of the model depends only upon the previous states of model

Markov Assumption

- Today's weather can always be predicted solely given knowledge of the weather of the past few days
- Unrealistic – But simplifies the analysis

Weather Example

States: S_1: rain; S_2: cloudy; S_3: sunny



Markov Process

Markov process is a process which moves from state to state depending (only) on the previous *n* states.

- *Order n* model
  - *n* is the number of states affecting the choice of next state

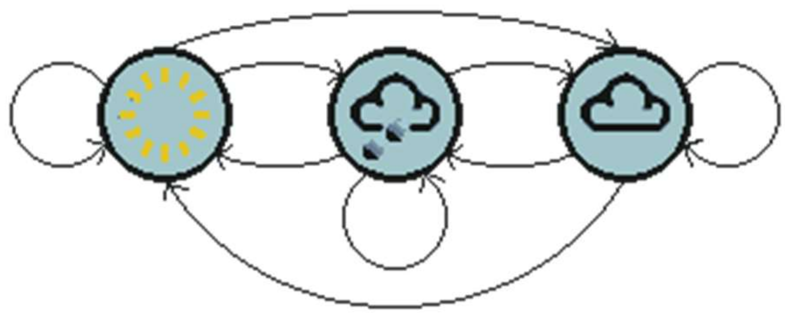
Markov Process

- Simplest Markov process
  - First order process where the choice of state is made purely on the basis of the previous state

How is it different from Deterministic???

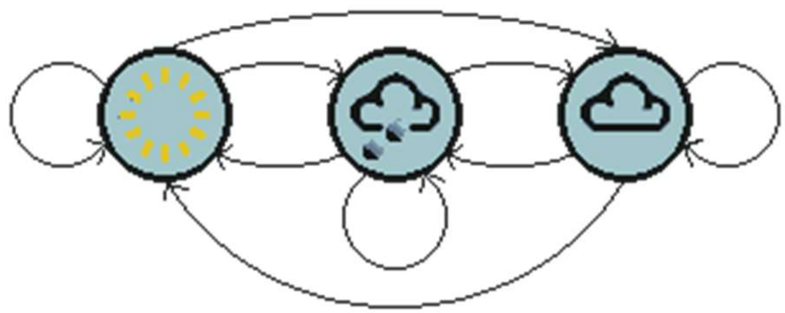
Choice to be made probabalistically, not deterministically

Markov Process



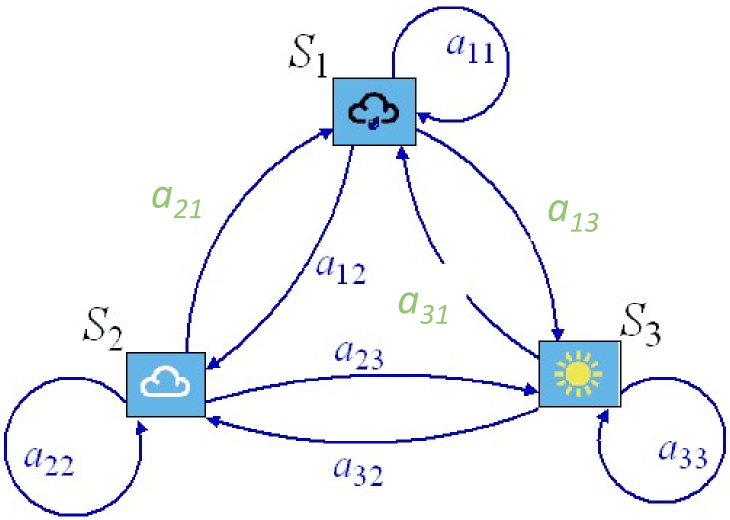
- M states - M^2 transitions between states
- State transition probability
  - Probability of moving from one state to another
- Collected in state transition matrix

Markov Process



$$\begin{array}{cc} & Today \\ Yesterday & \begin{array}{l} \\ \text{sun} \\ \text{cloud} \\ \text{rain} \end{array} \begin{bmatrix} \text{sun} & \text{cloud} & \text{rain} \\ 0.50 & 0.375 & 0.125 \\ 0.25 & 0.125 & 0.625 \\ 0.25 & 0.375 & 0.375 \end{bmatrix} \end{array}$$

Markov Process



3-state Markov model

$A=\{a_{i,j}\}$

$$\begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.6 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$$

Transitional probabilities $a_{i,j}$

Markov Process

- Initialization
  - Vector of initial probabilities π

$$
\begin{array}{ccc}
\text{Sun} & \text{Cloud} & \text{Rain} \\
\end{array}
$$
$$
\begin{pmatrix} 1.0 & 0.0 & 0.0 \end{pmatrix}
$$

Markov Model

- We have defined a first order Markov model comprising
  - **States** : Three states - sunny, cloudy, rainy
  - π **vector** : Defining the probability of the system being in each of the states at time 0
  - **State transition matrix** : The probability of the weather given the previous day's weather

Markov Model

- Set of states:

$$\{s_1, s_2, \ldots, s_N\}$$

- Process moves from one state to another generating a sequence of states :

$$s_{i1}, s_{i2}, \ldots, s_{ik}, \ldots$$

Markov Model

- Markov chain property: probability of each subsequent state depends only on what was the previous state:

$$P(s_{ik} \mid s_{i1}, s_{i2}, \ldots, s_{ik-1}) = P(s_{ik} \mid s_{ik-1})$$

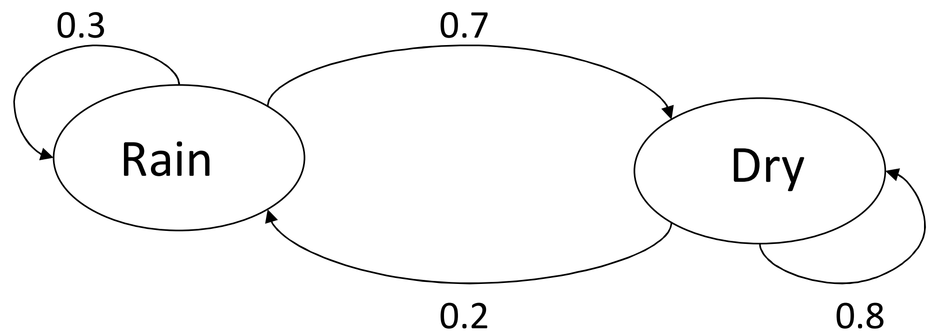
- To define Markov model, the following probabilities have to be specified:
  - Transition probabilities

$$a_{ij} = P(s_j \mid s_i)$$

  - Initial probabilities

$$\pi_i = P(s_i)$$

Markov Model – Example



- Two states : ‘Rain’ and ‘Dry’
- Transition probabilities:
  P(‘Rain’|‘Rain’)=0.3
  P(‘Dry’|‘Rain’)= 0.7
  P(‘Rain’|‘Dry’)= 0.2
  P(‘Dry’|‘Dry’)= 0.8
- Initial probabilities:
- P(‘Rain’)=0.4 P(‘Dry’)= 0.6

Calculation of Sequence Probability

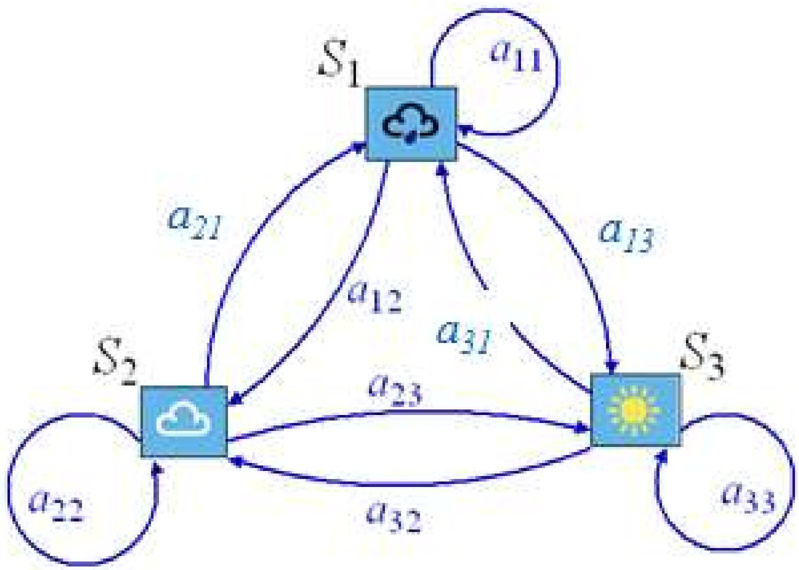
- Problem:
  - Calculate the probability of a sequence of states: {‘Dry’,’Dry’,’Rain’,Rain’}

P({‘Dry’,’Dry’,’Rain’,Rain’}) =

P(‘Rain’|’Rain’) P(‘Rain’|’Dry’) P(‘Dry’|’Dry’) P(‘Dry’) =

0.3*0.2*0.8*0.6

Calculation of Sequence Probability



State at time *k* is *i*:

Notation 1

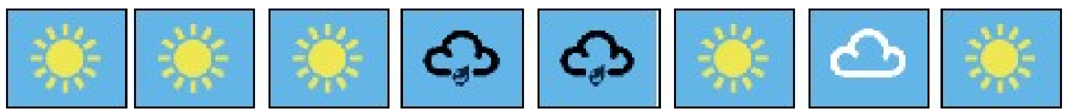
S_{ik}

Notation 2

$q_k = S_i$

Calculation of Sequence Probability

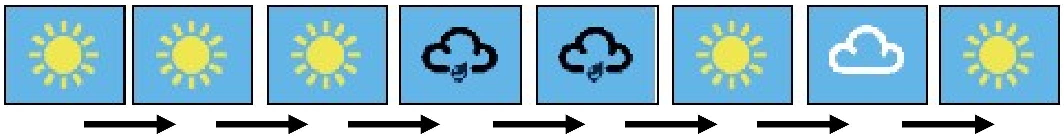
- On day t=1 the weather is sunny: q_1=S_3.
  What is the probability that the weather for the next 7 days will be "sun-sun-rain-rain-sun-cloudy-sun"?



t=1 t=2 t=3 t=4 t=5 t=6 t=7 t=8

Observation sequence: O=S_3,S_3, S_3, S_1, S_1, S_3, S_2, S_3;

Calculation of Sequence Probability



$a_{3,3}$ $a_{3,3}$ $a_{3,1}$ $a_{1,1}$ $a_{1,3}$ $a_{3,2}$ $a_{2,3}$

$$P(O)=P\{S_3,\ S_{3,}S_3,\ S_1,\ S_1,\ S_3,\ S_2,\ S_3\}=$$

$$= P(q_1)\cdot\ P(S_3|S_3)\cdot\ P(S_3|S_3)\cdot\ P(S_1|S_3)\cdot P(S_1|S_1)\cdot P(S_3|S_1)\ \cdot P(S_2|S_3)\cdot P(S_3|S_2)=$$

$$=1\cdot\ a_{3,3}\ \cdot a_{3,3}\cdot a_{3,1}\cdot a_{1,1}\cdot a_{1,3}\cdot a_{3,2}\cdot a_{2,3}=$$

$$=1\cdot(0.8)\cdot(0.8)\cdot(0.1)\cdot(0.4)\cdot(0.3)\cdot(0.1)\cdot(0.2)=1.54\cdot 10^{-4}$$

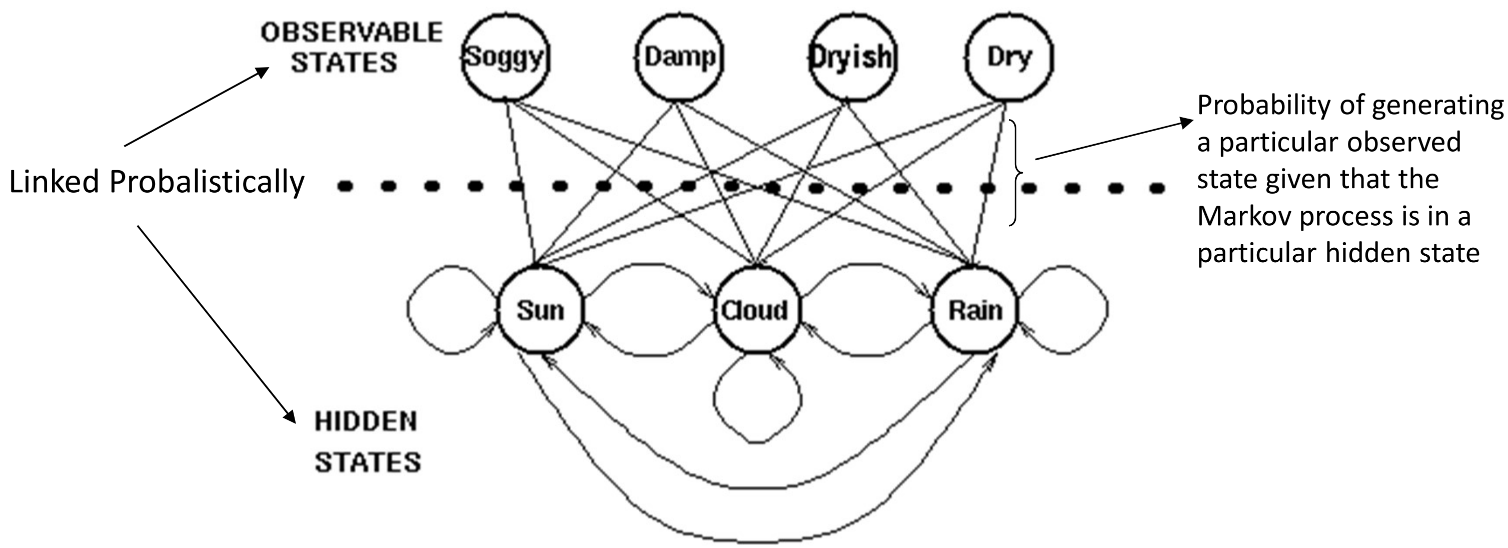
Hidden Markov Models

Example

- You are locked in a room for several days
- You are asked about the weather outside
- Only evidence:
  - Care taker is carrying an umbrella or not
  - Piece of seaweed
- Evidence is somehow linked to the weather

Observed States – Hidden States???

Hidden Markov Models



Assumption: Hidden states (the true weather) are modeled by a simple first order Markov process

Hidden Markov Models

Matrix of observation probabilities

- Contains the probabilities of the observable states given a particular hidden state

$$
\begin{array}{ccc}
 & & \text{Seaweed} \\
 & & \begin{array}{cccc} \textbf{Dry} & \textbf{Dryish} & \textbf{Damp} & \textbf{Soggy} \end{array} \\
\text{weather} & \begin{array}{c} \textbf{Sun} \\ \textbf{Cloud} \\ \textbf{Rain} \end{array} & \begin{pmatrix} 0.60 & 0.20 & 0.15 & 0.05 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0.05 & 0.10 & 0.35 & 0.50 \end{pmatrix}
\end{array}
$$

Hidden Markov Models

- Set of states:

$$\{s_1, s_2, \ldots, s_N\}$$

- Process moves from one state to another generating a sequence of states :

$$s_{i1}, s_{i2}, \ldots, s_{ik}, \ldots$$

- Markov chain property: probability of each subsequent state depends only on what was the previous state:

$$P(s_{ik} \mid s_{i1}, s_{i2}, \ldots, s_{ik-1}) = P(s_{ik} \mid s_{ik-1})$$

Hidden Markov Models

- States are not visible, but each state randomly generates one of M observations (or visible states)

$$\{v_1, v_2, \ldots, v_M\}$$

Hidden Markov Models

Defining HMM

- To define hidden Markov model, the following probabilities have to be specified:
  - Matrix of transition probabilities $A=(a_{ij})$
  - Matrix of observation probabilities $B=(b_i(v_m))$,
    $b_i(v_m) = P(v_m | s_i)$
  - Vector of initial probabilities $\pi=(\pi_i)$, $\pi_i = P(s_i)$.
  - Model is represented by $M=(A, B, \pi)$

HMM – Three Problems

Question # 1 – Evaluation

GIVEN

A sequence of observations:

Dry Dry Dry Rainy Rainy Rainy Dry Rainy

QUESTION

How likely is this sequence, given our model

This is the **EVALUATION** problem in HMMs

HMM – Three Problems

- Evaluation Problem

Given the HMM $M=(A, B, \pi)$ and the observation sequence $O=o_1\ o_2 \ldots o_K$, calculate the probability that model M has generated sequence O

$O=o_1 \ldots o_K$ denotes a sequence of observations $o_k \in \{v_1, \ldots, v_M\}$.

HMM – Forward Algorithm Example

Model Description

- Hidden States
  - Sunny, Cloudy, Rainy
- Observable States
  - Dry, Dryish, Damp, Soggy
- Initial State Probabilities

	Sunny	Cloudy	Rainy
π	0.63	0.17	0.20

HMM – Forward Algorithm Example

- State Transition Matrix A

	weather today		
weather yesterday	Sunny	Cloudy	Rainy
Sunny	0.500	0.375	0.125
Cloudy	0.250	0.125	0.625
Rainy	0.250	0.375	0.375

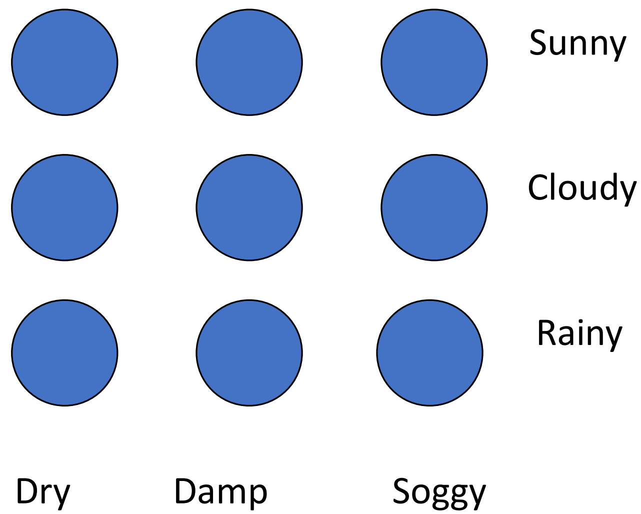
- Observational Probability Matrix B

	observed states			
hidden states	Dry	Dryish	Damp	Soggy
Sunny	0.60	0.20	0.15	0.05
Cloudy	0.25	0.25	0.25	0.25
Rainy	0.05	0.10	0.35	0.50

	Sunny	Cloudy	Rainy
π	0.63	0.17	0.20

	weather today		
weather yesterday	**Sunny**	**Cloudy**	**Rainy**
Sunny	0.500	0.375	0.125
Cloudy	0.250	0.125	0.625
Rainy	0.250	0.375	0.375

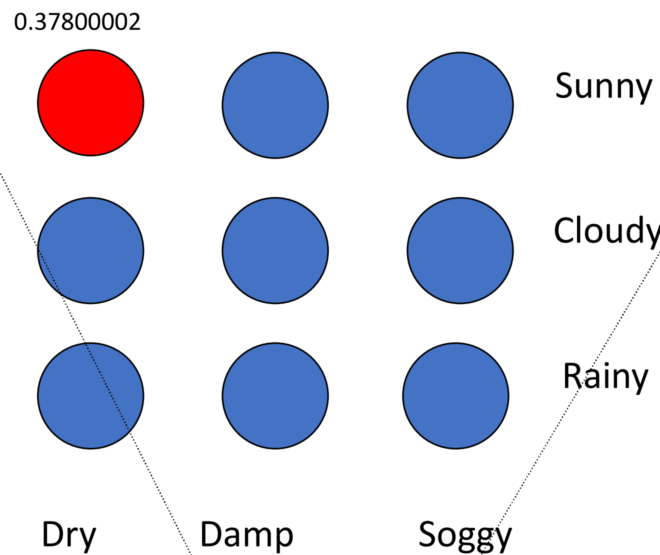
	observed states			
hidden states	**Dry**	**Dryish**	**Damp**	**Soggy**
Sunny	0.60	0.20	0.15	0.05
Cloudy	0.25	0.25	0.25	0.25
Rainy	0.05	0.10	0.35	0.50



	Sunny	Cloudy	Rainy
π	0.63	0.17	0.20

weather yesterday	weather today: Sunny	Cloudy	Rainy
Sunny	0.500	0.375	0.125
Cloudy	0.250	0.125	0.625
Rainy	0.250	0.375	0.375

hidden states	observed states: Dry	Dryish	Damp	Soggy
Sunny	0.60	0.20	0.15	0.05
Cloudy	0.25	0.25	0.25	0.25
Rainy	0.05	0.10	0.35	0.50

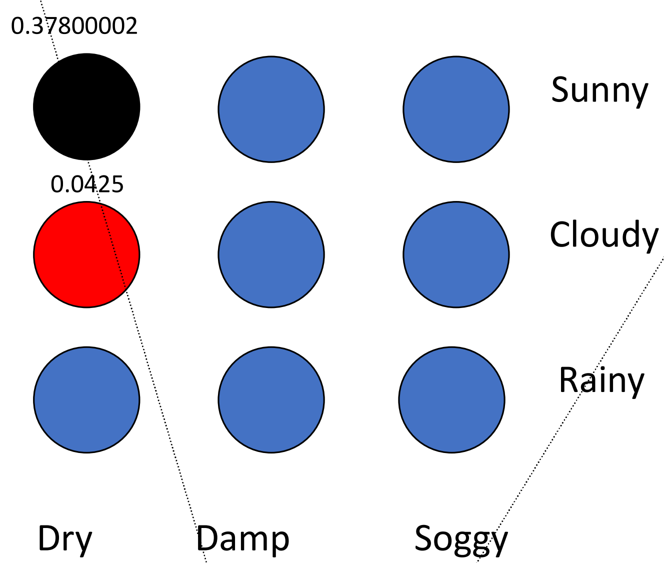


$\text{Alpha}_1(\text{Sunny}) = (0.63 * 0.6) = 0.37800002$

	Sunny	Cloudy	Rainy
π	0.63	0.17	0.20

weather yesterday	weather today		
	Sunny	Cloudy	Rainy
Sunny	0.500	0.375	0.125
Cloudy	0.250	0.125	0.625
Rainy	0.250	0.375	0.375

hidden states	observed states			
	Dry	Dryish	Damp	Soggy
Sunny	0.60	0.20	0.15	0.05
Cloudy	0.25	0.25	0.25	0.25
Rainy	0.05	0.10	0.35	0.50

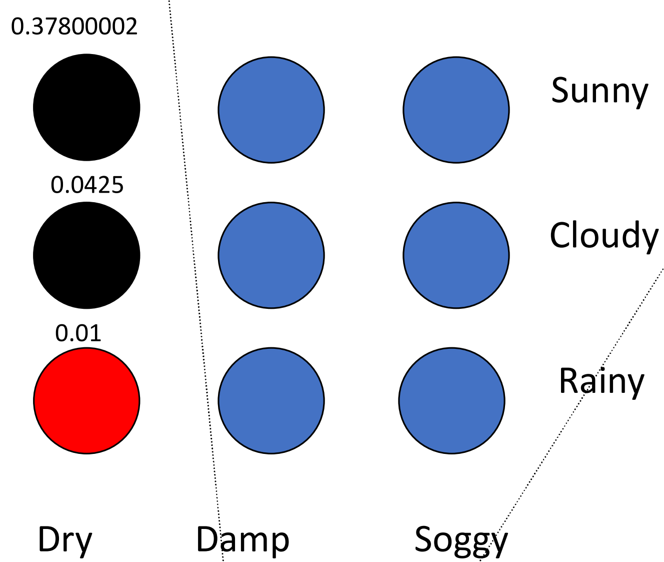


Alpha$_1$(Cloudy) = (0.17 * 0.25) = 0.0425

	Sunny	Cloudy	Rainy
π	0.63	0.17	0.20

		weather today	
weather yesterday	Sunny	Cloudy	Rainy
Sunny	0.500	0.375	0.125
Cloudy	0.250	0.125	0.625
Rainy	0.250	0.375	0.375

		observed states		
hidden states	Dry	Dryish	Damp	Soggy
Sunny	0.60	0.20	0.15	0.05
Cloudy	0.25	0.25	0.25	0.25
Rainy	0.05	0.10	0.35	0.50

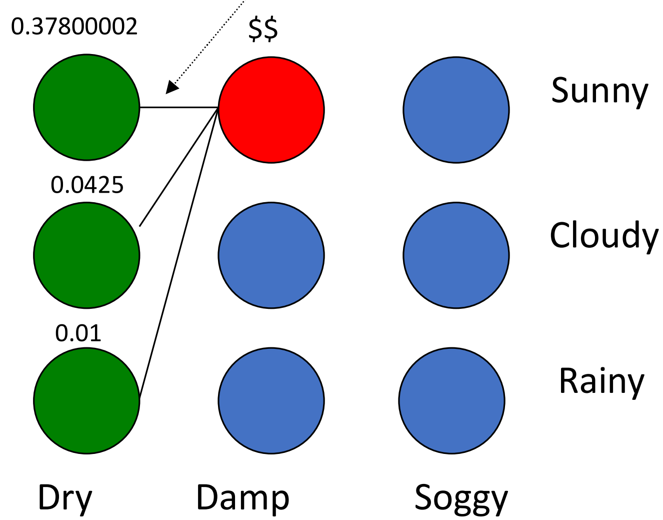


$\text{Alpha}_1(\text{Rainy}) = (0.2 * 0.05) = 0.010000001$

	Sunny	Cloudy	Rainy
π	0.63	0.17	0.20

weather yesterday	weather today: Sunny	Cloudy	Rainy
Sunny	0.500	0.375	0.125
Cloudy	0.250	0.125	0.625
Rainy	0.250	0.375	0.375

hidden states	observed states: Dry	Dryish	Damp	Soggy
Sunny	0.60	0.20	0.15	0.05
Cloudy	0.25	0.25	0.25	0.25
Rainy	0.05	0.10	0.35	0.50



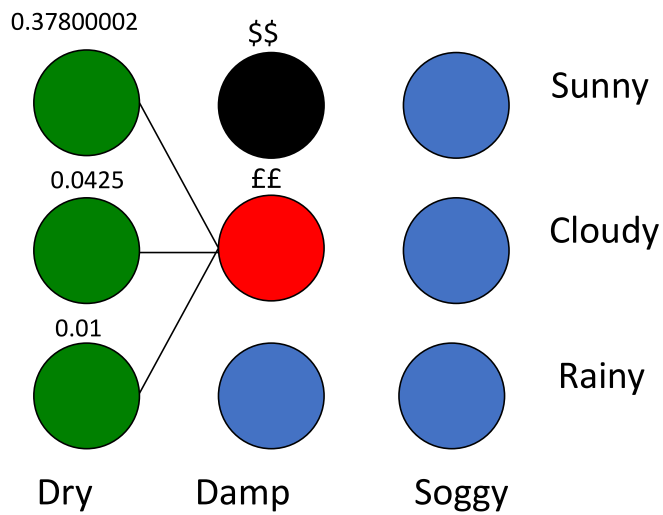
$\text{Alpha}_2(\text{Sunny}) = [\text{Alpha}_1(\text{Sunny}) * a_{ss} + \text{Alpha}_1(\text{Cloudy}) * a_{cs} + \text{Alpha}_1(\text{Rainy}) * a_{rs}] * b_{s,damp}$

Alpha2(Sunny) = (((0.37800002*0.5) + (0.0425*0.25) + (0.010000001*0. 25)) * 0.15)=$$

	Sunny	Cloudy	Rainy
π	0.63	0.17	0.20

		weather today		
		Sunny	Cloudy	Rainy
weather yesterday	Sunny	0.500	0.375	0.125
	Cloudy	0.250	0.125	0.625
	Rainy	0.250	0.375	0.375

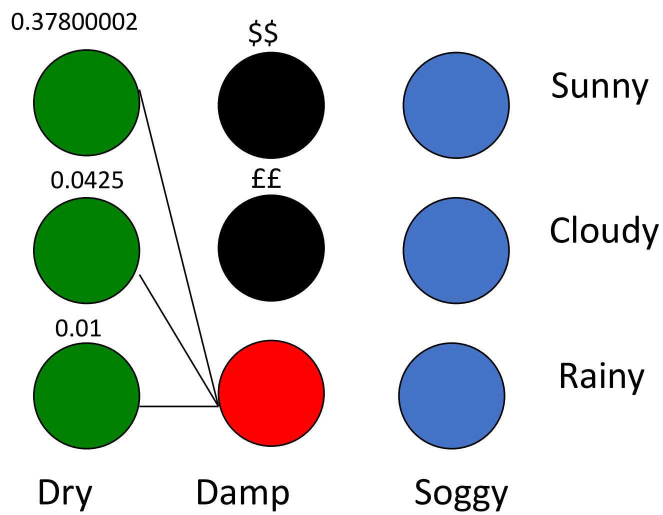
		observed states			
		Dry	Dryish	Damp	Soggy
hidden states	Sunny	0.60	0.20	0.15	0.05
	Cloudy	0.25	0.25	0.25	0.25
	Rainy	0.05	0.10	0.35	0.50



	Sunny	Cloudy	Rainy
π	0.63	0.17	0.20

weather yesterday	weather today		
	Sunny	**Cloudy**	**Rainy**
Sunny	0.500	0.375	0.125
Cloudy	0.250	0.125	0.625
Rainy	0.250	0.375	0.375

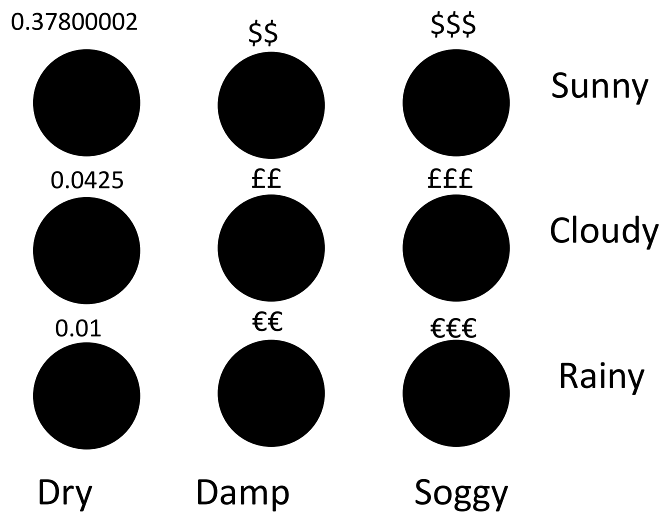
hidden states	observed states			
	Dry	**Dryish**	**Damp**	**Soggy**
Sunny	0.60	0.20	0.15	0.05
Cloudy	0.25	0.25	0.25	0.25
Rainy	0.05	0.10	0.35	0.50



	Sunny	Cloudy	Rainy
π	0.63	0.17	0.20

weather yesterday	weather today: Sunny	Cloudy	Rainy
Sunny	0.500	0.375	0.125
Cloudy	0.250	0.125	0.625
Rainy	0.250	0.375	0.375

hidden states	observed states: Dry	Dryish	Damp	Soggy
Sunny	0.60	0.20	0.15	0.05
Cloudy	0.25	0.25	0.25	0.25
Rainy	0.05	0.10	0.35	0.50



Probability of the model = $$$+£££+€€€

HMM – Three Problems

Question # 2 – Decoding

GIVEN

A sequence of observations:

Dry Dry Dry Rainy Rainy Rainy Dry Rainy

QUESTION

Which portion of the sequence was generated by 'low' atmospheric pressure and what portion corresponds to 'high' atmospheric pressure

This is the **DECODING** problem in HMMs

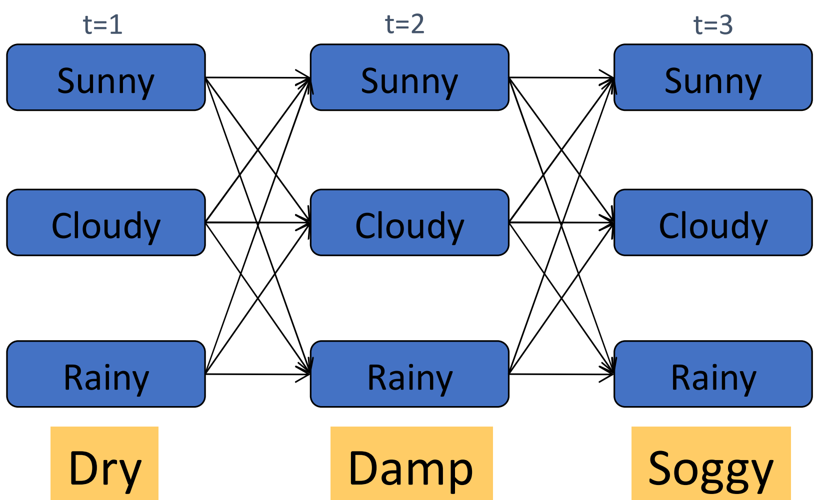
HMM – Three Problems

- Decoding Problem

Given the HMM $M=(A, B, \pi)$ and the observation sequence $O=o_1\ o_2\ ...\ o_K$, calculate the most likely sequence of hidden states s_i that produced this observation sequence O.

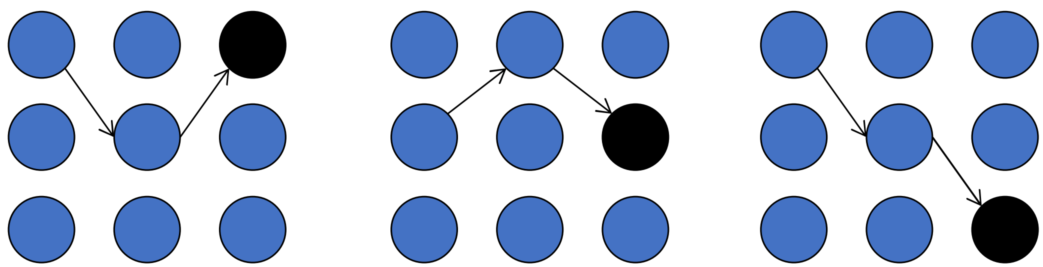
HMM – Viterbi Algorithm

- Finding the most probable sequence of hidden states



For each intermediate and terminating state in the trellis there is a most probable path to that state

HMM – Viterbi Alogrithm



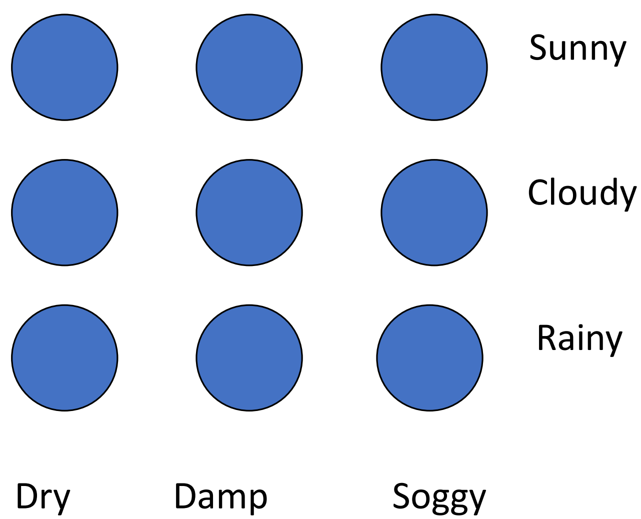
Partial best paths each having a probability δ

Unlike the partial probabilities in the forward algorithm, δ is the probability of the one (most probable) path to the state.

	Sunny	Cloudy	Rainy
π	0.63	0.17	0.20

weather yesterday	weather today		
	Sunny	Cloudy	Rainy
Sunny	0.500	0.250	0.250
Cloudy	0.375	0.125	0.375
Rainy	0.125	0.675	0.375

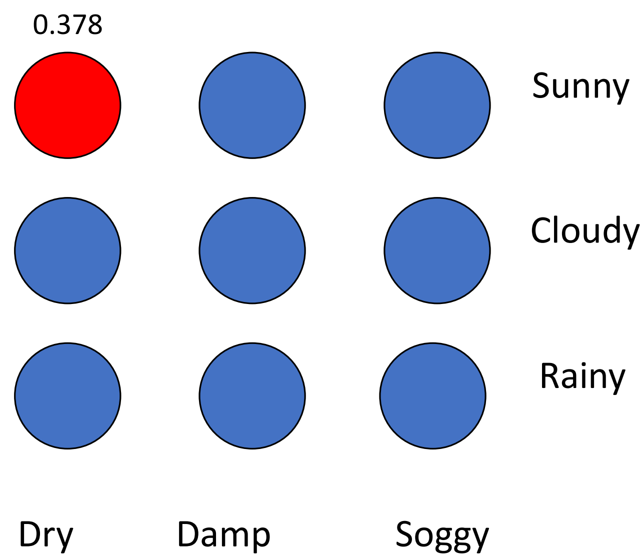
hidden states	observed states			
	Dry	Dryish	Damp	Soggy
Sunny	0.60	0.20	0.15	0.05
Cloudy	0.25	0.25	0.25	0.25
Rainy	0.05	0.10	0.35	0.50



	Sunny	Cloudy	Rainy
π	0.63	0.17	0.20

weather yesterday	weather today: Sunny	Cloudy	Rainy
Sunny	0.500	0.250	0.250
Cloudy	0.375	0.125	0.375
Rainy	0.125	0.675	0.375

hidden states	observed states: Dry	Dryish	Damp	Soggy
Sunny	0.60	0.20	0.15	0.05
Cloudy	0.25	0.25	0.25	0.25
Rainy	0.05	0.10	0.35	0.50

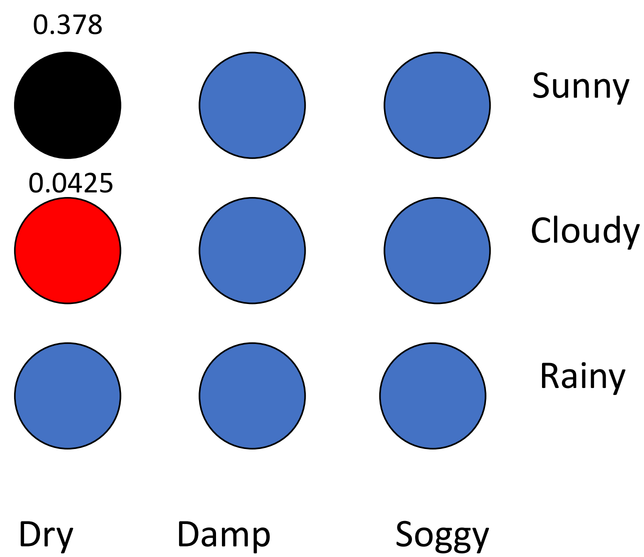


$\delta_1(\text{Sunny}) = 0.63 * 0.6 = 0.37800002$

	Sunny	Cloudy	Rainy
π	0.63	0.17	0.20

weather yesterday	weather today		
	Sunny	Cloudy	Rainy
Sunny	0.500	0.250	0.250
Cloudy	0.375	0.125	0.375
Rainy	0.125	0.675	0.375

hidden states	observed states			
	Dry	Dryish	Damp	Soggy
Sunny	0.60	0.20	0.15	0.05
Cloudy	0.25	0.25	0.25	0.25
Rainy	0.05	0.10	0.35	0.50

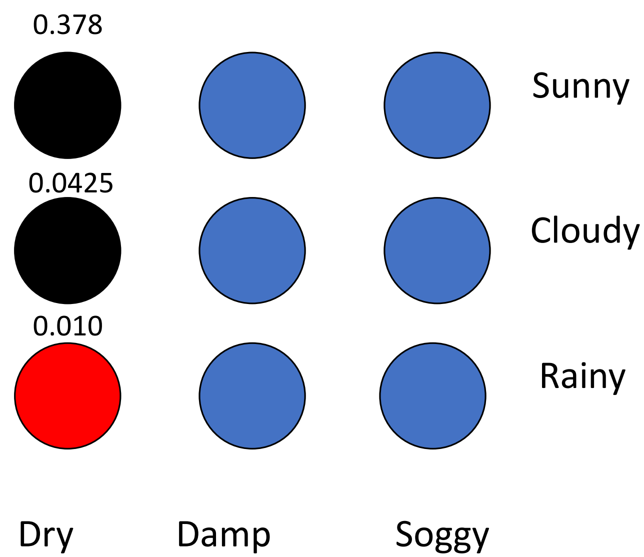


δ_1(Cloudy)=0.17 * 0.25 = 0.0425

	Sunny	Cloudy	Rainy
π	0.63	0.17	0.20

weather yesterday	weather today		
	Sunny	Cloudy	Rainy
Sunny	0.500	0.250	0.250
Cloudy	0.375	0.125	0.375
Rainy	0.125	0.675	0.375

hidden states	observed states			
	Dry	Dryish	Damp	Soggy
Sunny	0.60	0.20	0.15	0.05
Cloudy	0.25	0.25	0.25	0.25
Rainy	0.05	0.10	0.35	0.50

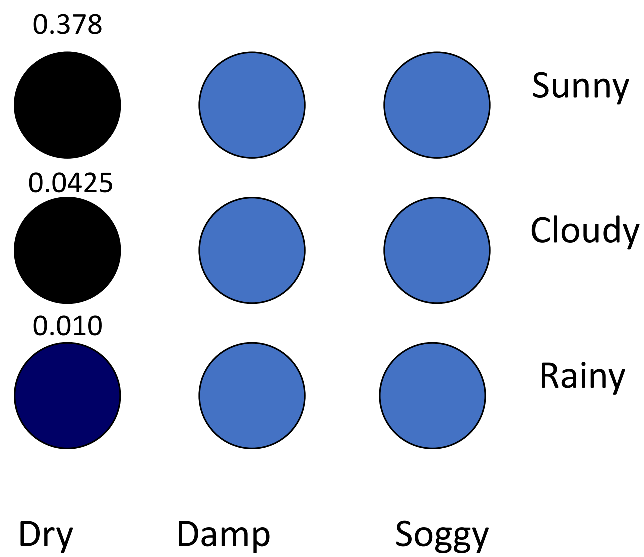


δ_1(Rainy)=0.2 * 0.05 = 0.010000001

	Sunny	Cloudy	Rainy
π	0.63	0.17	0.20

weather yesterday	weather today		
	Sunny	Cloudy	Rainy
Sunny	0.500	0.250	0.250
Cloudy	0.375	0.125	0.375
Rainy	0.125	0.675	0.375

hidden states	observed states			
	Dry	Dryish	Damp	Soggy
Sunny	0.60	0.20	0.15	0.05
Cloudy	0.25	0.25	0.25	0.25
Rainy	0.05	0.10	0.35	0.50

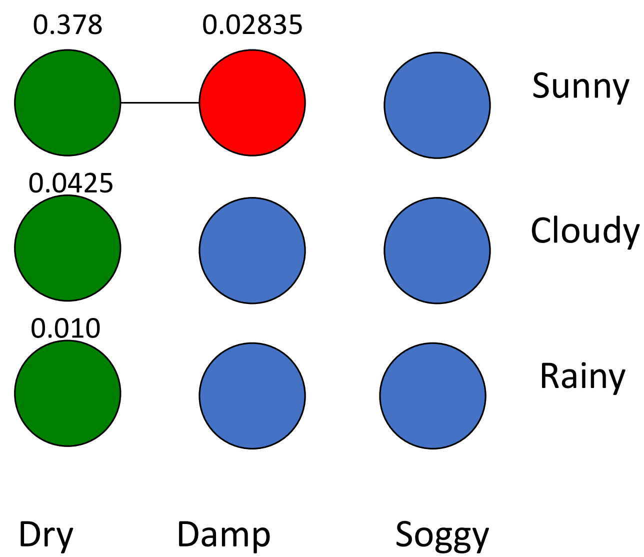


$\delta_2(\text{Sunny}) = ?$

	Sunny	Cloudy	Rainy
π	0.63	0.17	0.20

	weather today		
weather yesterday	Sunny	Cloudy	Rainy
Sunny	0.500	0.250	0.250
Cloudy	0.375	0.125	0.375
Rainy	0.125	0.675	0.375

	observed states			
hidden states	Dry	Dryish	Damp	Soggy
Sunny	0.60	0.20	0.15	0.05
Cloudy	0.25	0.25	0.25	0.25
Rainy	0.05	0.10	0.35	0.50



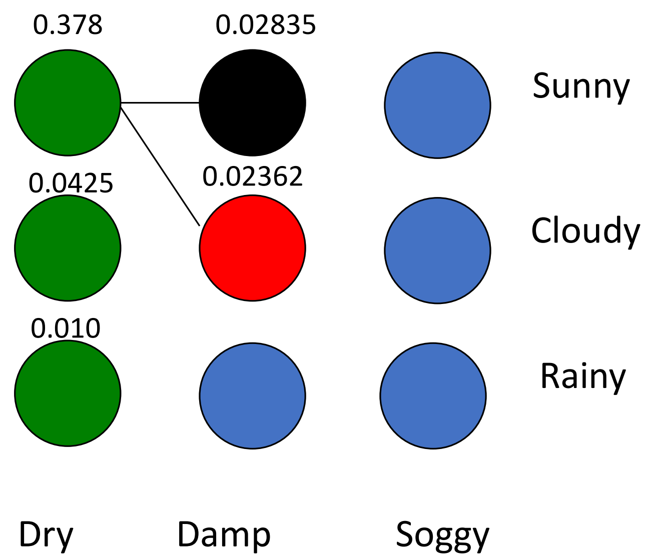
δ_2(Sunny)=max ((0.37800002*0.5), (0.0425*0.375), (0.010000001*0.125)) * 0.15 = 0.028350003

Φ_2(Sunny) = Sunny

	Sunny	Cloudy	Rainy
π	0.63	0.17	0.20

weather yesterday		weather today	
	Sunny	Cloudy	Rainy
Sunny	0.500	0.250	0.250
Cloudy	0.375	0.125	0.375
Rainy	0.125	0.675	0.375

hidden states	observed states			
	Dry	Dryish	Damp	Soggy
Sunny	0.60	0.20	0.15	0.05
Cloudy	0.25	0.25	0.25	0.25
Rainy	0.05	0.10	0.35	0.50



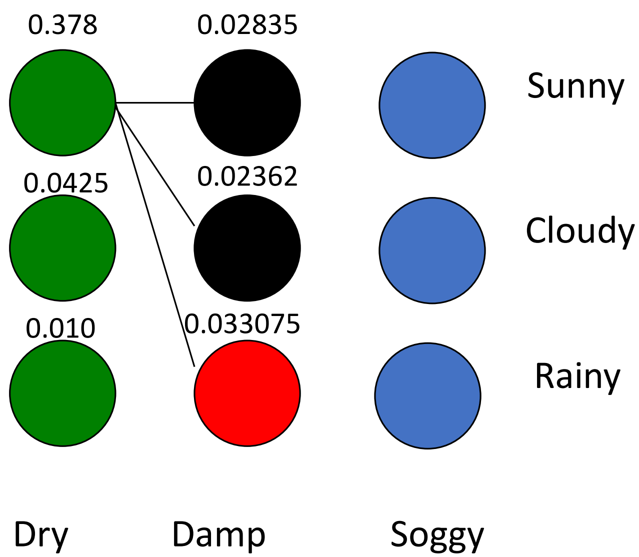
δ_2(Cloudy)=Delta = max ((0.37800002*0.25), (0.0425*0.125), (0.010000001*0.675)) * 0.25 = 0.023625001

Φ_2(Cloudy) = Sunny

	Sunny	Cloudy	Rainy
π	0.63	0.17	0.20

		weather today		
		Sunny	Cloudy	Rainy
weather yesterday	Sunny	0.500	0.250	0.250
	Cloudy	0.375	0.125	0.375
	Rainy	0.125	0.675	0.375

		observed states			
		Dry	Dryish	Damp	Soggy
hidden states	Sunny	0.60	0.20	0.15	0.05
	Cloudy	0.25	0.25	0.25	0.25
	Rainy	0.05	0.10	0.35	0.50



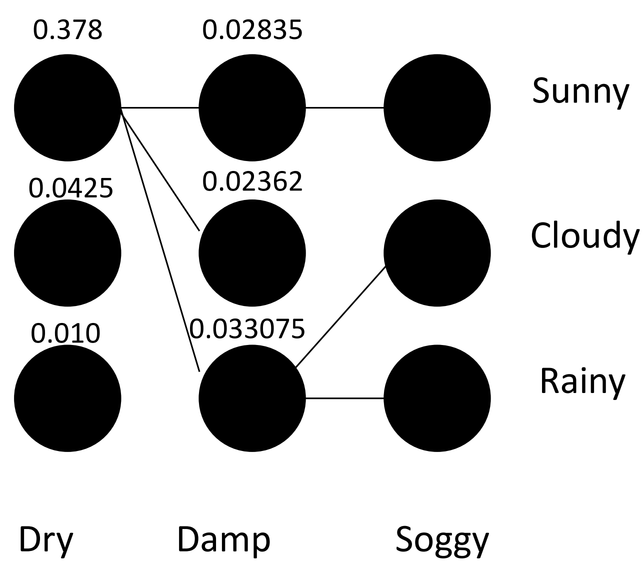
δ_2(Cloudy)=max ((0.37800002*0.25), (0.0425*0.375), (0.010000001*0.375)) * 0.35 = 0.033075

$\Phi_2(Rainy) = Sunny$

	Sunny	Cloudy	Rainy
π	0.63	0.17	0.20

weather yesterday	weather today		
	Sunny	Cloudy	Rainy
Sunny	0.500	0.250	0.250
Cloudy	0.375	0.125	0.375
Rainy	0.125	0.675	0.375

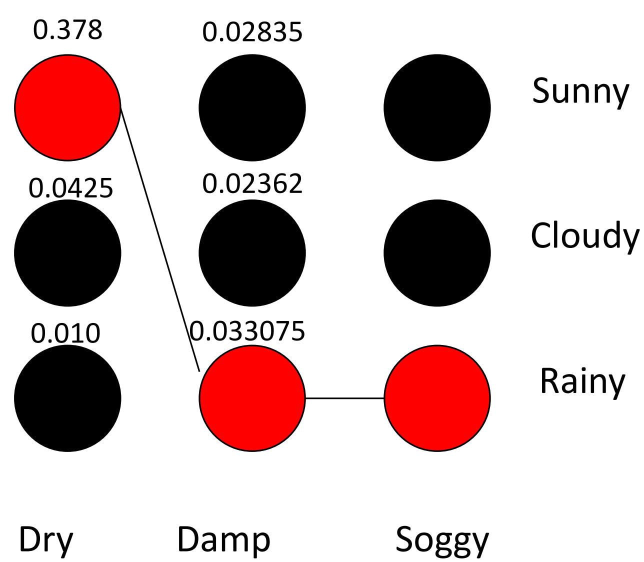
hidden states	observed states			
	Dry	Dryish	Damp	Soggy
Sunny	0.60	0.20	0.15	0.05
Cloudy	0.25	0.25	0.25	0.25
Rainy	0.05	0.10	0.35	0.50



	Sunny	Cloudy	Rainy
π	0.63	0.17	0.20

		weather today		
		Sunny	Cloudy	Rainy
weather yesterday	Sunny	0.500	0.250	0.250
	Cloudy	0.375	0.125	0.375
	Rainy	0.125	0.675	0.375

		observed states			
		Dry	Dryish	Damp	Soggy
hidden states	Sunny	0.60	0.20	0.15	0.05
	Cloudy	0.25	0.25	0.25	0.25
	Rainy	0.05	0.10	0.35	0.50



HMM – Three Problems

Question # 3 – Learning

GIVEN

Sequence(s) of observations:

Dry Dry Dry Rainy Rainy Rainy Dry Rainy

And possible hidden states

Low Pressure, High Pressure

QUESTION

How the hidden states are linked to the observable states? How often pressure changes from low to high etc.

This is the **LEARNING** problem in HMMs

HMM – Three Problems

- Learning Problem

Given some training observation sequences $O=o_1\ o_2\ ...\ o_K$ and general structure of HMM (numbers of hidden and visible states), determine HMM parameters $M=(A, B, \pi)$ that best fit training data

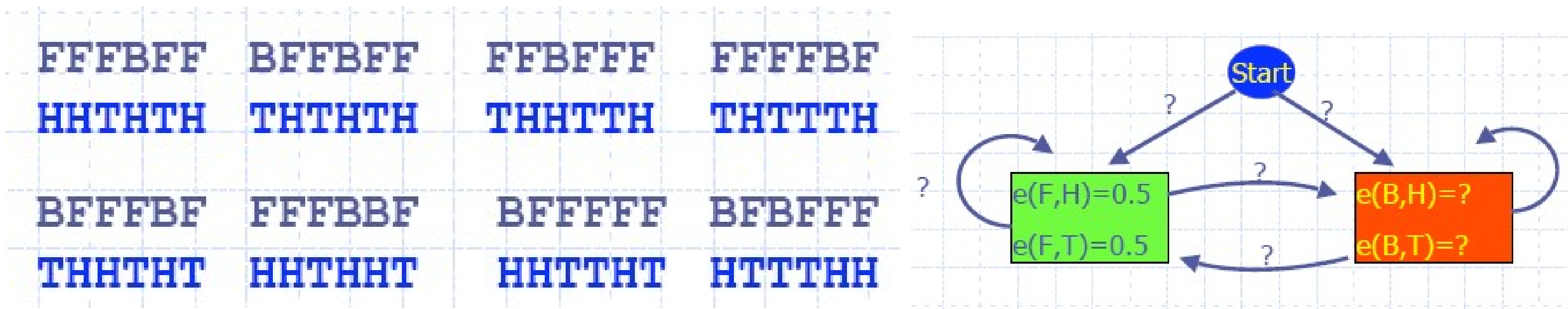
Supervised Training

If training data has information about sequence of hidden states then use maximum likelihood estimation of parameters:

$$a_{ji} = P(s_i \mid s_j) = \frac{\text{Number of transitions from state } S_j \text{ to state } S_i}{\text{Number of transitions out of state } S_j}$$

$$b_i(v_m) = P(v_m \mid s_i) = \frac{\text{Number of times observation } V_m \text{ occurs in state } S_i}{\text{Number of times in state } S_i}$$

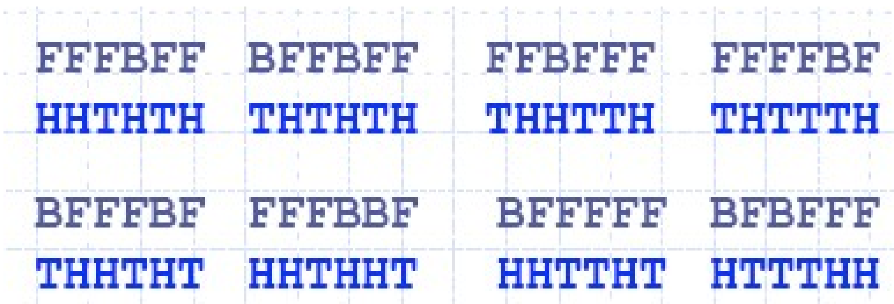
Supervised Training

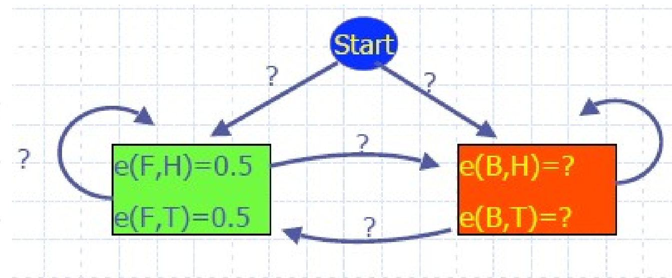


- Count transition frequencies
  - Start $\longrightarrow$ Fair = 4
  - π_{fair} = 4/8 = 0.5
  - Fair $\longrightarrow$ Biased = 7
  - a(Fair, Biased) = 7/(4+3+4+4+3+3+4+3) = ¼

{F⇒B} transitions occur at 7 locations ($1_3,2_3,3_2,4_4,5_4,6_3,8_2$) where k_i denotes location i of sequence k

Supervised Training





- Observation/Emission Frequencies
  - E(Biased, Head) = 8
  - e(Biased, Head) = 8/12

The material in these slides is based on the following resources

References

- A tutorial on Hidden Markov Models by Roger Boyle, University of Leeds
- Chapter 13, Introduction to Machine Learning, E. Alpyadin, MIT Press
- Hidden Markov Models by V. Govindaraju
- Hidden Markov Models by Y. Yemini, Columbia University