Deterministic Finite State Transducers

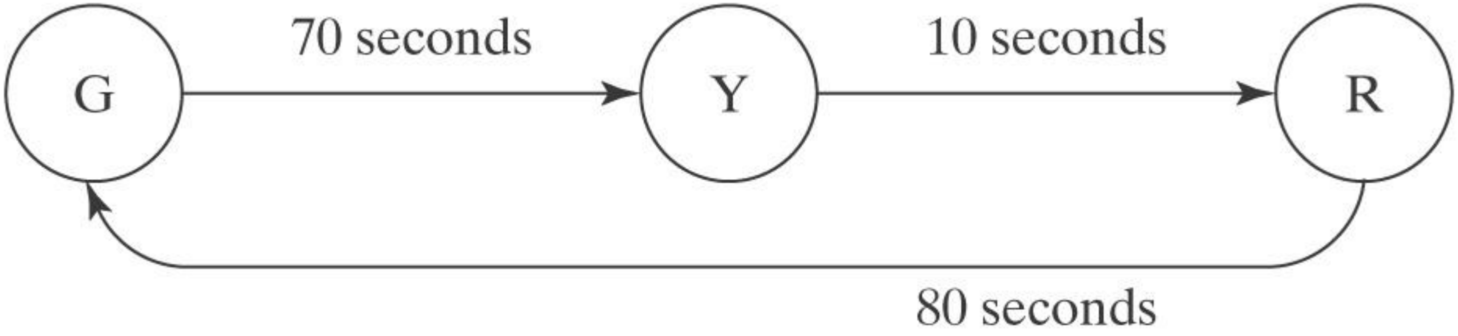
A ***Moore machine*** $M = (K, \Sigma, O, \delta, D, s, A)$, where:

- K is a finite set of states
- Σ is an input alphabet
- O is an output alphabet
- $s \in K$ is the initial state
- $A \subseteq K$ is the set of accepting states,
- δ is the transition function from $(K \times \Sigma)$ to (K),
- D is the output function from (K) to (O^*).

M outputs each time it lands in a state.

A Moore machine M computes a function $f(w)$ iff, when it reads the input string w, its output sequence is $f(w)$.

A Simple US Traffic Light Controller



Deterministic Finite State Transducers

A ***Mealy machine*** $M = (K, \Sigma, O, \delta, s, A)$, where:

- K is a finite set of states
- Σ is an input alphabet
- O is an output alphabet
- $s \in K$ is the initial state
- $A \subseteq K$ is the set of accepting states
- δ is the transition function from $(K \times \Sigma)$ to $(K \times O^*)$

- M outputs each time it takes a transition.

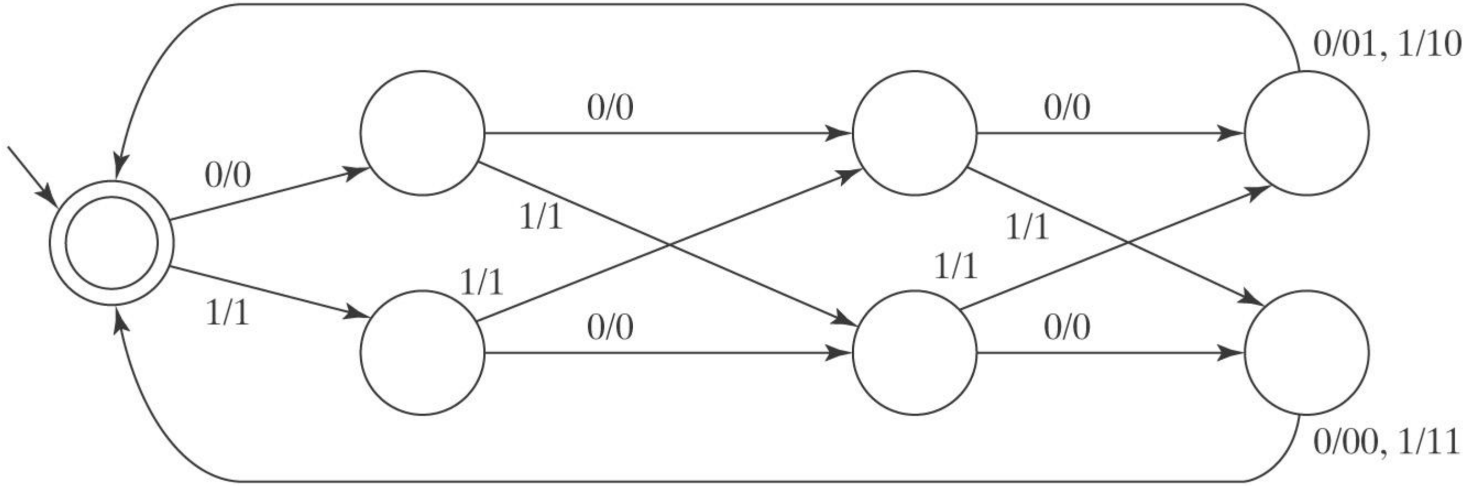
- A Mealy machine M computes a function $f(w)$ iff, when it reads the input string w, its output sequence is $f(w)$.

Deterministic Finite State Transducers

An Odd Parity Generator :

After every four bits, output a fifth bit such that each group of four bits has odd parity.

0 0 1 0   1 1 0 0   0 0 0 0   1 1 1 1



A Deterministic Finite State Transducer Interpreter

Let:

δ_1(*state*, *symbol*) return a single new state, and

δ_2(*state*, *symbol*) return an element of O^*.

$ST := s$.

Repeat

i := get-next-symbol.

If $i \neq$ end-of-file then:

Write($\delta_2(ST, i)$).

$ST := \delta_1(ST, i)$.

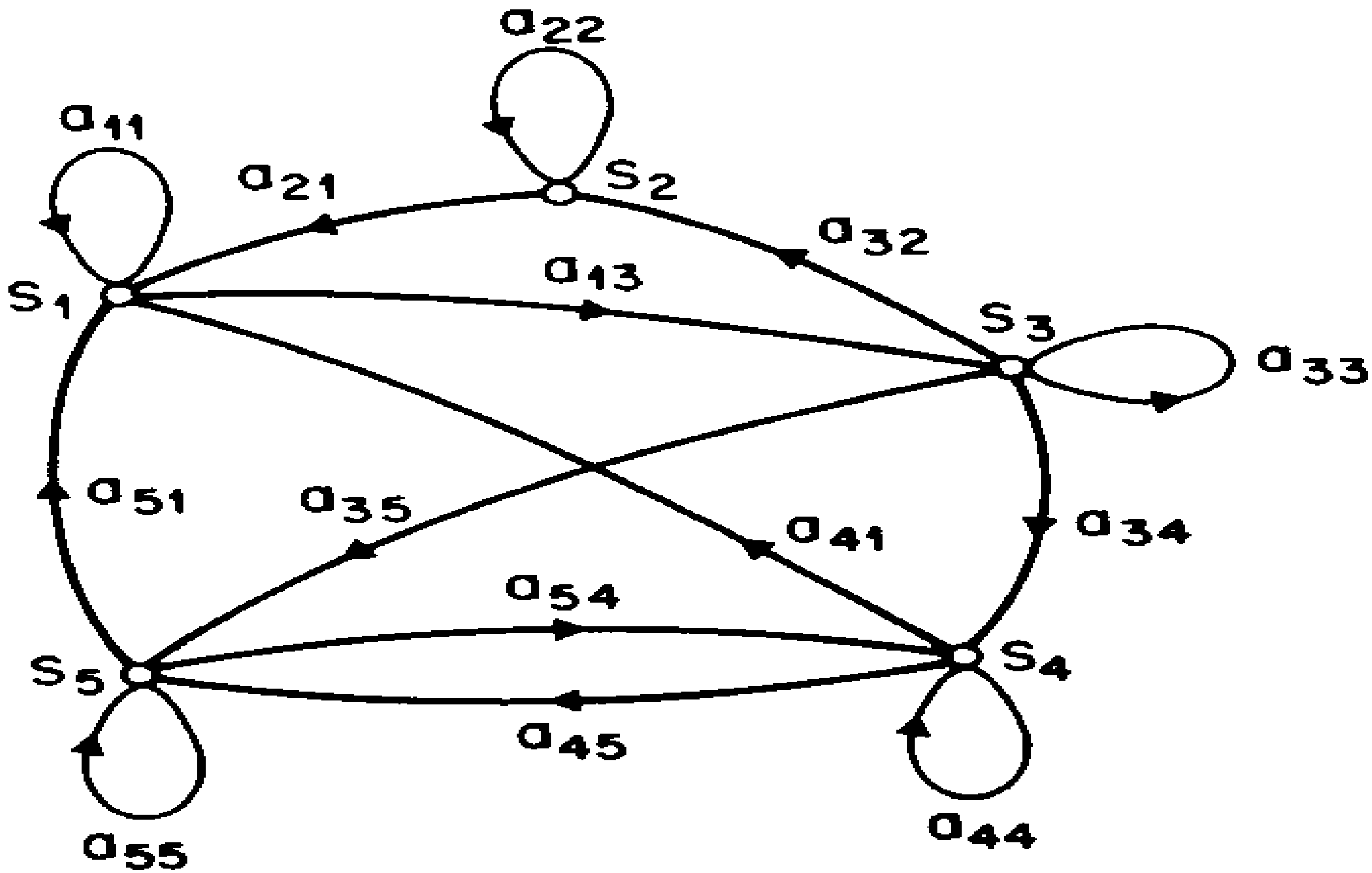
Until i = end-of-file.

Markov Models

A Markov model is a triple $M = (K, \pi, A)$:

- K is a finite set of states
- π is a vector of initial probabilities
- $A[p, q]$ = Pr(state q at time t | state p at t - 1)

Discrete Markov Process



A system with five states

Discrete Markov Process

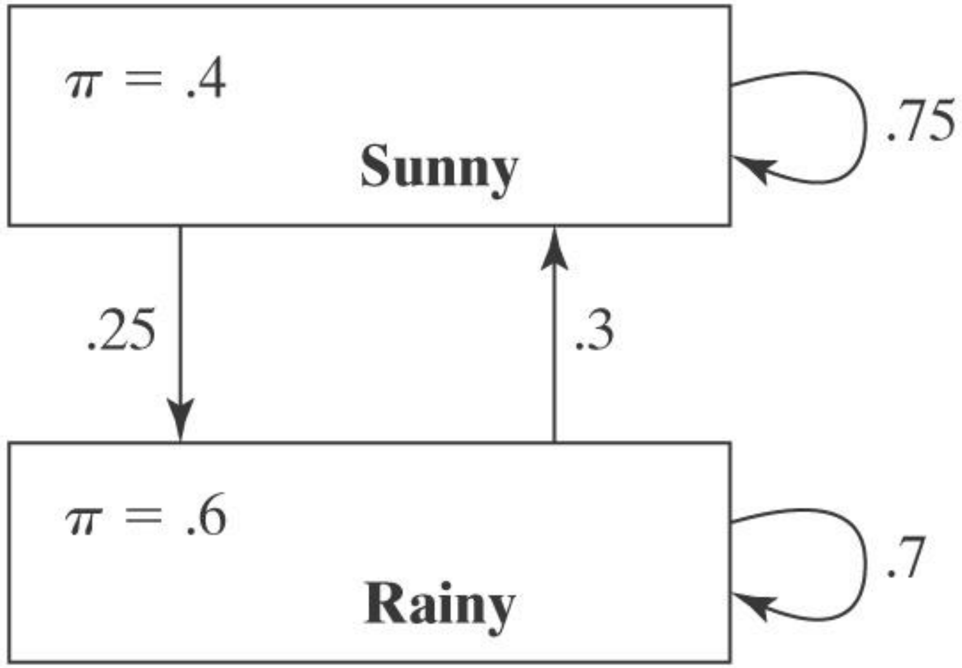
- The system undergoes changes at regular time intervals t=1,2,3 …
- The probability of the system going from state s_i to state s_j is denoted by a_{ij}.
- In a first order Markov Chain the state at time t depends only on the state at time t-1.
- a_{ij}'s are state transition coefficients

$$a_{ij} = P[q_t = S_j | q_{t-1} = S_i], \quad 1 \leq i, j \leq N$$

$$a_{ij} \geq 0$$

$$\sum_{j=1}^{N} a_{ij} = 1$$

Markov Models



What is the probability that it will be sunny five days in a row?

$.4 \cdot (.75)^4 = .1266$

Markov Models

To use a Markov model, we first need to use data to create the matrix *A*.

What can we do with a Markov model?

- Generate almost natural behavior.
- Estimate the probability of some outcome.

Markov Models

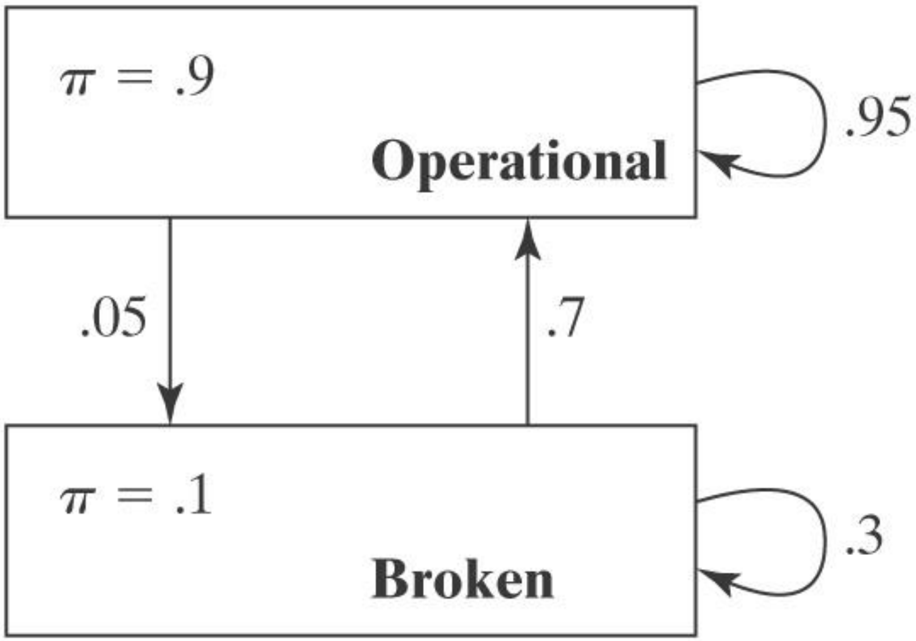
Estimating Probabilities

Given a Markov model that describes some random process, what is the probability that we will observe a particular sequence $S_1\ S_2\ \ldots\ S_n$ of states?

$$\Pr(s_1 s_2 \ldots s_n) = \pi[s_1] \cdot \prod_{i=2}^{n} A[s_{i-1}, s_i]$$

Markov Models

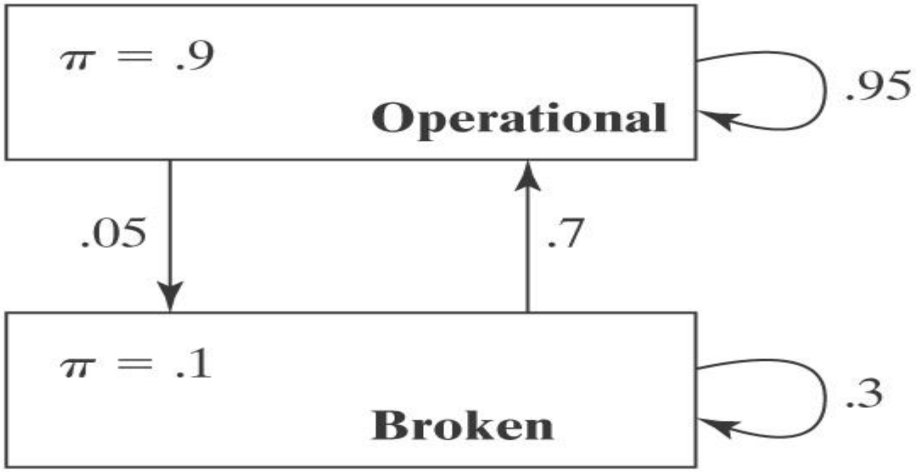
Modeling System Performance



If up now, what is probability of staying up for an hour?

Markov Models

Modeling System Performance



If up now, what is probability of staying up for an hour (3600 time steps)?

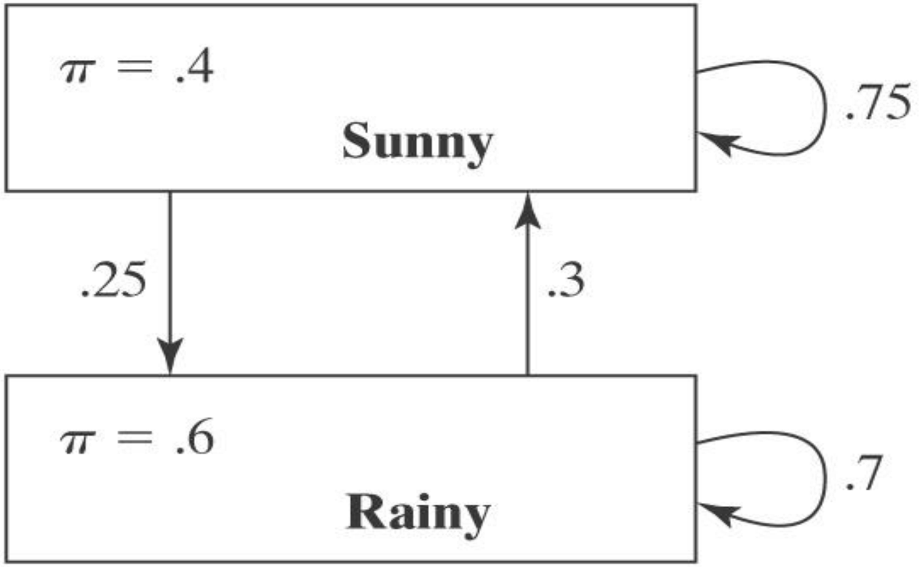
$$\Pr(s_1 s_2 \ldots s_n) = \pi[s_1] \cdot \prod_{i=2}^{n} A[s_{i-1}, s_i] = .95^{3600} = 6.3823 \cdot 10^{-81}$$

Hidden Markov Models

http://www.cs.ubc.ca/~murphyk/Bayes/rabiner.pdf

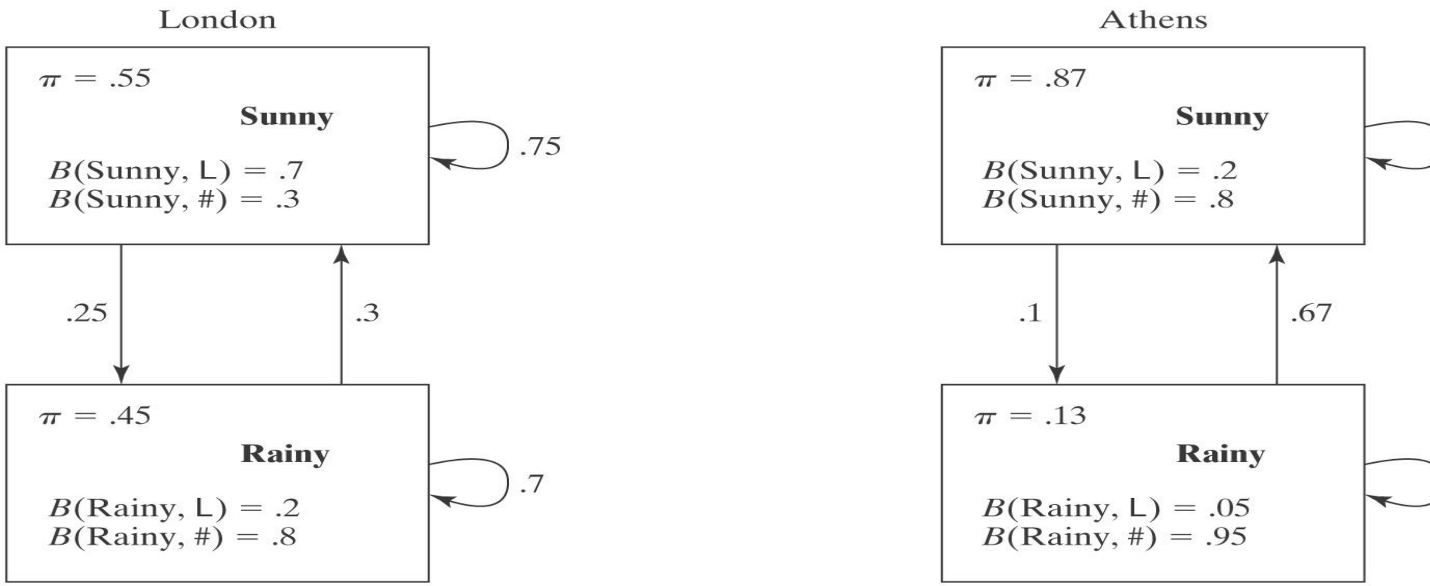
Suppose that the states themselves are not visible. But states emit outputs with certain probabilities and the outputs are visible

If we could view the states:



Hidden Markov Models

If we could not view the states:



- The evaluation problem: We observe the report ###L from somewhere.
- The decoding problem: We observe the report ###L from London.
- The learning problem.

Hidden Markov Models

Problem 1: Given the observation sequence $O = O_1 O_2 \cdots O_T$, and a model $\lambda = (A, B, \pi)$, how do we efficiently compute $P(O|\lambda)$, the probability of the observation sequence, given the model?

Problem 2: Given the observation sequence $O = O_1 O_2 \cdots O_T$, and the model λ, how do we choose a corresponding state sequence $Q = q_1 q_2 \cdots q_T$ which is optimal in some meaningful sense (i.e., best "explains" the observations)?

Problem 3: How do we adjust the model parameters $\lambda = (A, B, \pi)$ to maximize $P(O|\lambda)$?

Hidden Markov Models

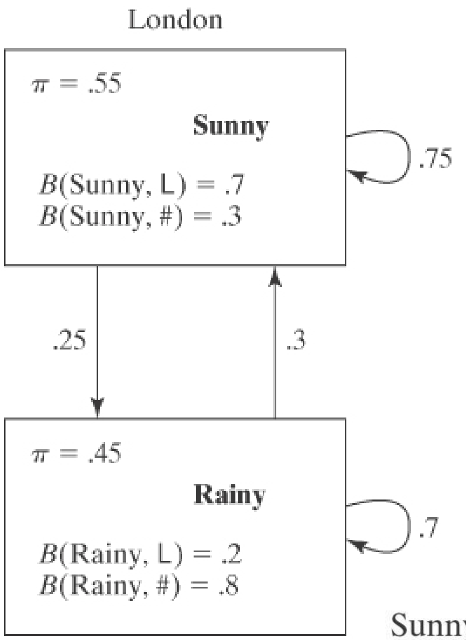
An HMM M is a quintuple (K, O, π, A, B), where:

- K is a finite set of states,
- O is the output alphabet,
- π is a vector of initial probabilities of the states,
- A is a matrix of transition probabilities:

$$A[p, q] = \Pr(\text{state } q \text{ at time } t \mid \text{state } p \text{ at time } t - 1),$$

- B, the confusion matrix of output probabilities.

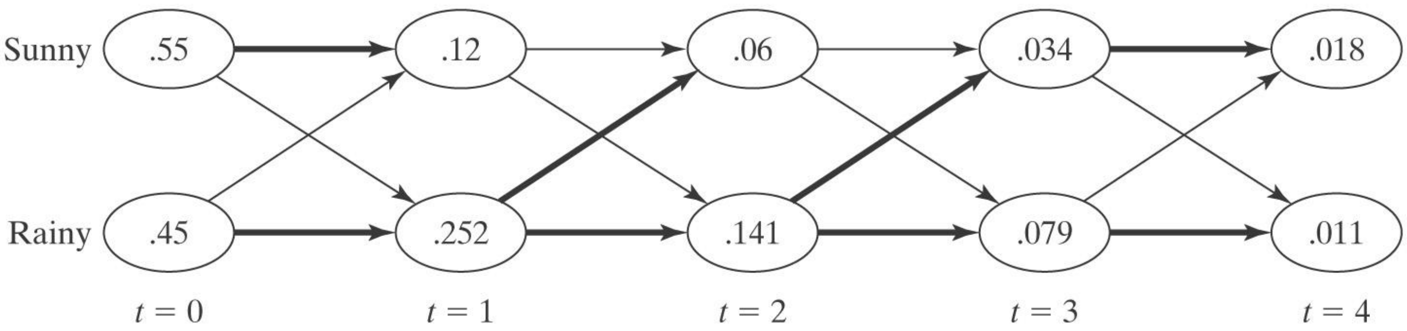
$$B[q, o] = \Pr(\text{output } o \mid \text{state } q).$$



The Decoding Problem

We observe the report ###L from London. What was the weather?

We use the Viterbi algorithm:



candidate-score[Sunny]	= *score*[Sunny, 0] · *A*[Sunny, Sunny] · *B*[Sunny, #]
	= .55 · .75 · .3
	= .12
candidate-score[Rainy]	= *score*[Rainy, 0] · *A*[Rainy, Sunny] · *B*[Rainy, #]
	= .45 · .3 · .8
	= .11

So *score*[Sunny, 1] = max(.12, .11) = .12, and *back*(Sunny, 1) is set to Sunny.