THE UNIVERSITY of TENNESSEE KNOXVILLE

Lecture 23: Markov Processes

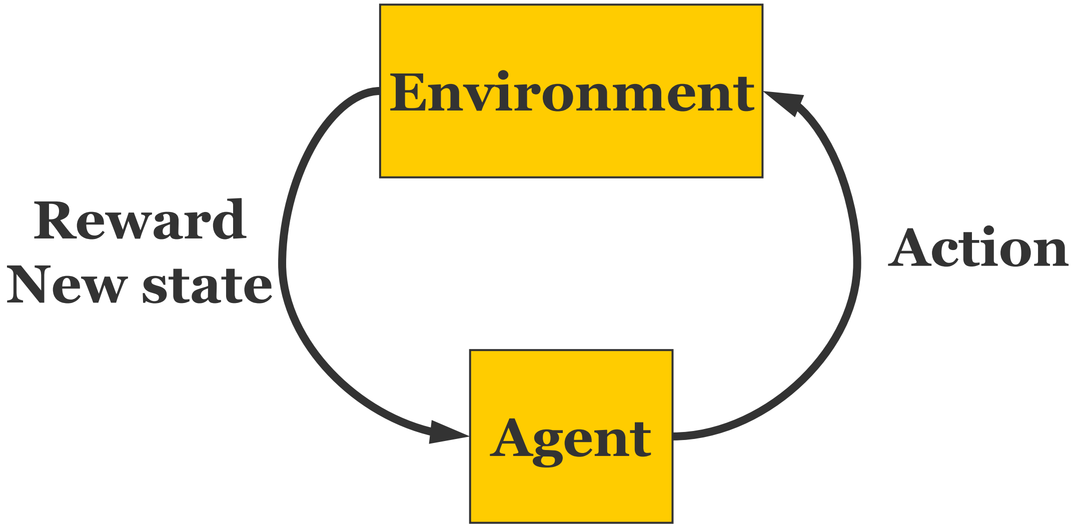
Instructor: Sergei V. Kalinin

(A Peek at) Reinforcement Learning

- **Supervised learning**
  – Classification
  – Regression

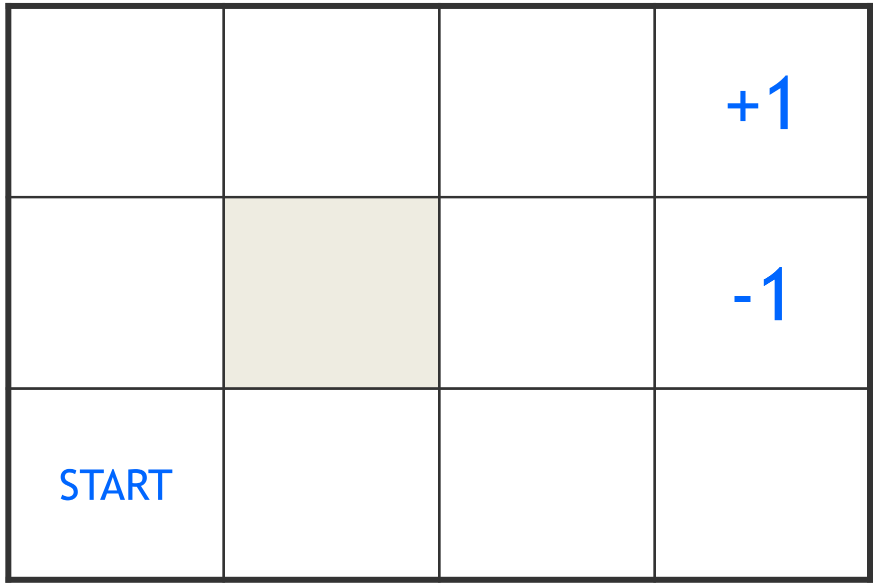
- **Unsupervised learning**
  – Clustering
  – Dimensionality reduction

- **Reinforcement learning**
  – more general than supervised/unsupervised learning
  – learn from interaction with environment to achieve a goal



From Reinforcement Learning Tutorial, Peter Bodík, RAD Lab, UC Berkeley

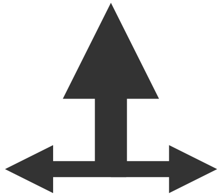
Robot in the room



Actions: UP, DOWN, LEFT, RIGHT

UP

80% move UP
10% move LEFT
10% move RIGHT

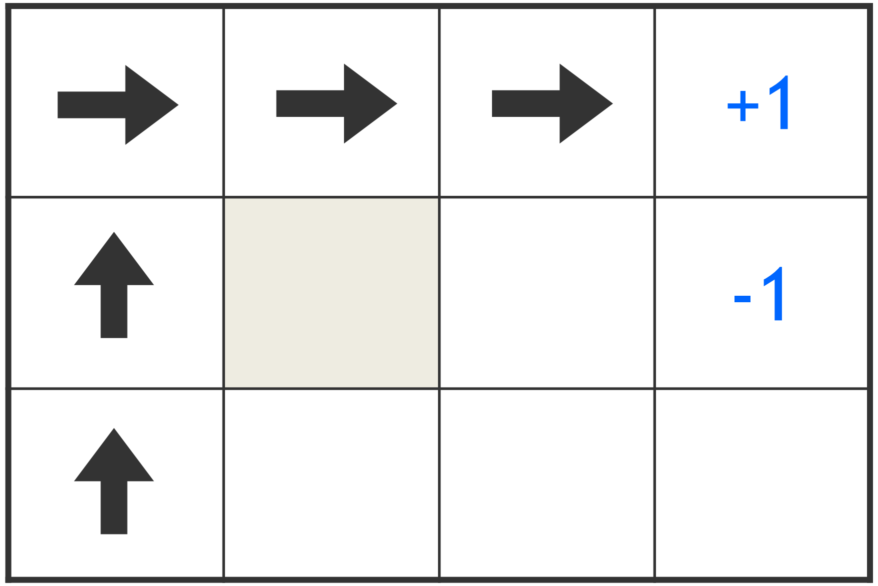


- Reward +1 at [4,3], -1 at [4,2]
- Reward -0.04 for each step

- What's the strategy to achieve max reward?
- What if the actions were deterministic?

From Reinforcement Learning Tutorial, Peter Bodík, RAD Lab, UC Berkeley

Is this a solution?

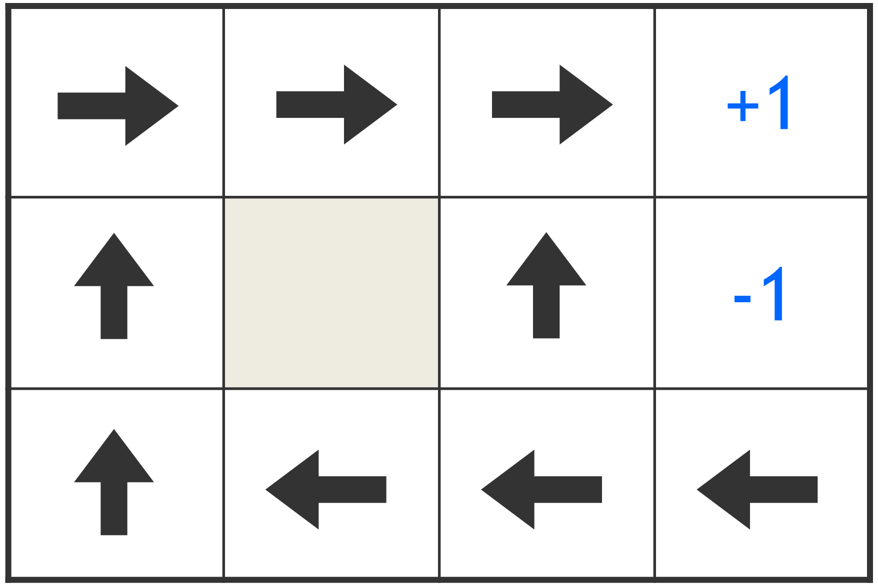


- only if actions deterministic
- not in this case (actions are stochastic)

- solution/policy
- mapping from each state to an action

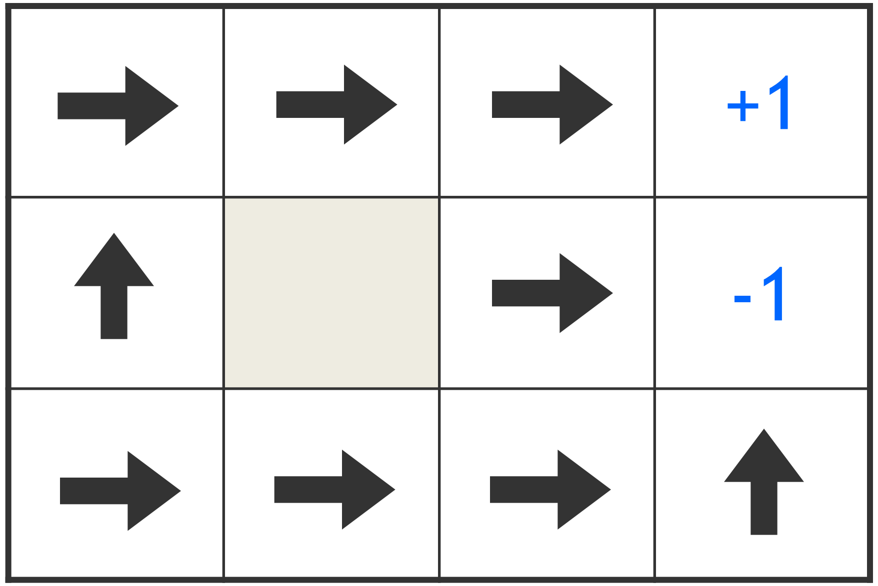
From Reinforcement Learning Tutorial, Peter Bodík, RAD Lab, UC Berkeley

Optimal policy



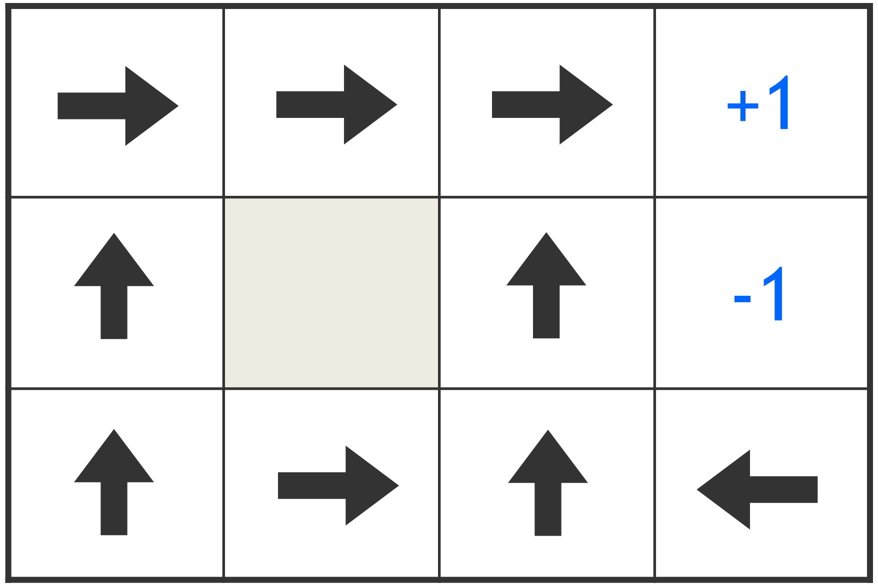
From Reinforcement Learning Tutorial, Peter Bodík, RAD Lab, UC Berkeley

What if the reward for each step is -2?



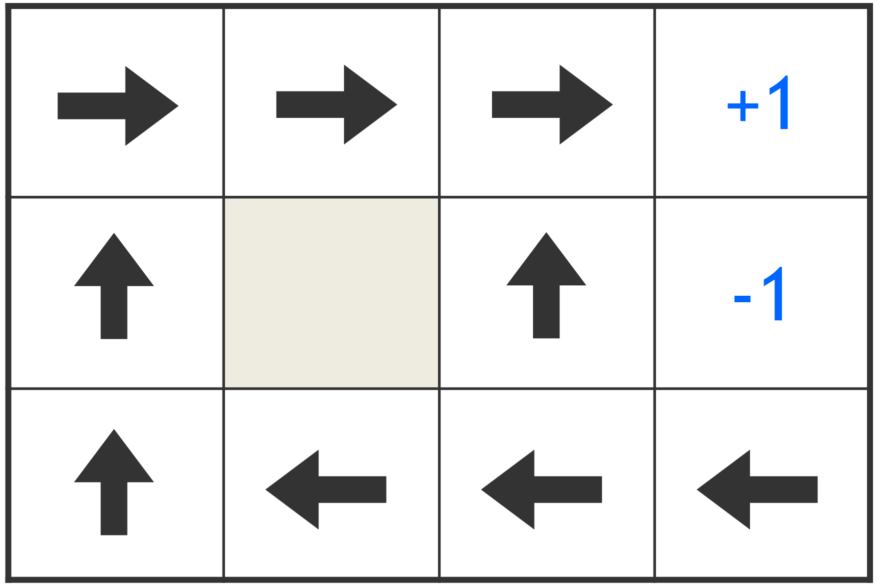
From Reinforcement Learning Tutorial, Peter Bodík, RAD Lab, UC Berkeley

Reward for each step is -0.1



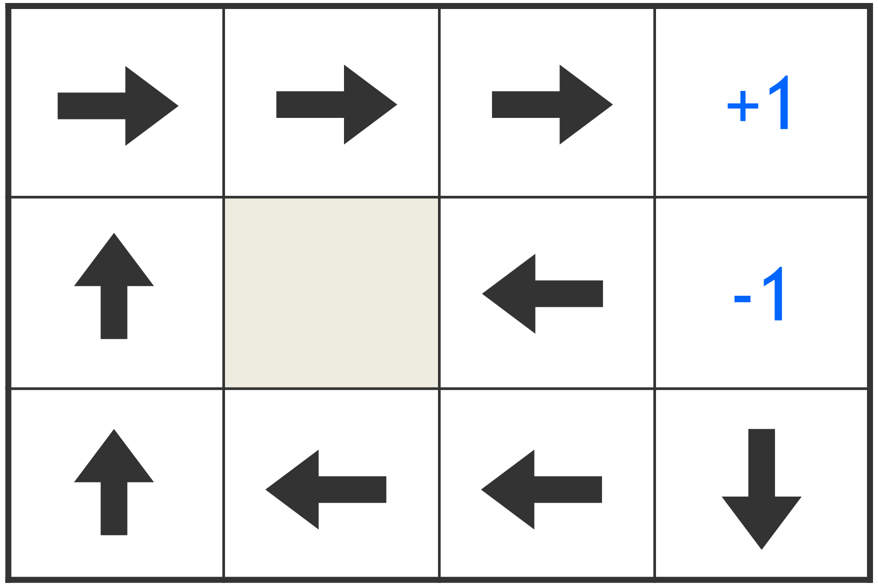
From Reinforcement Learning Tutorial, Peter Bodík, RAD Lab, UC Berkeley

Reward for each step is -0.04



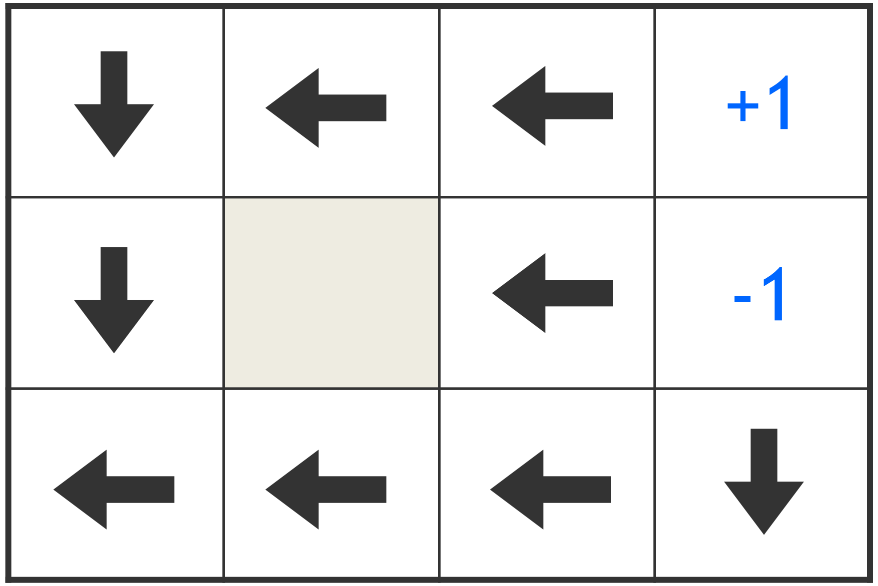
From Reinforcement Learning Tutorial, Peter Bodík, RAD Lab, UC Berkeley

Reward for each step is -0.01



From Reinforcement Learning Tutorial, Peter Bodík, RAD Lab, UC Berkeley

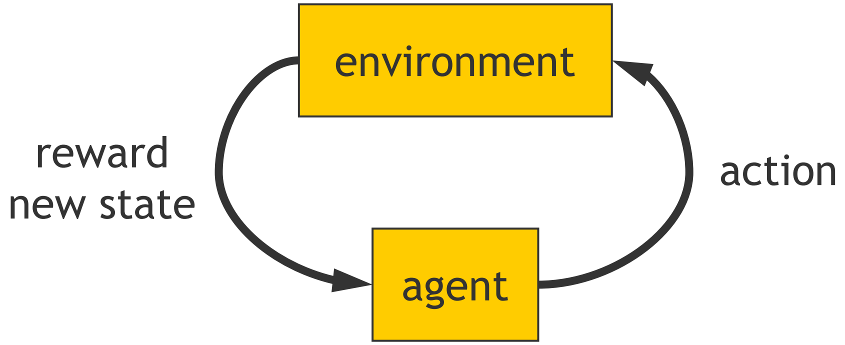
Reward for each step is +0.01



From Reinforcement Learning Tutorial, Peter Bodík, RAD Lab, UC Berkeley

Markov Decision Processes (MDP)

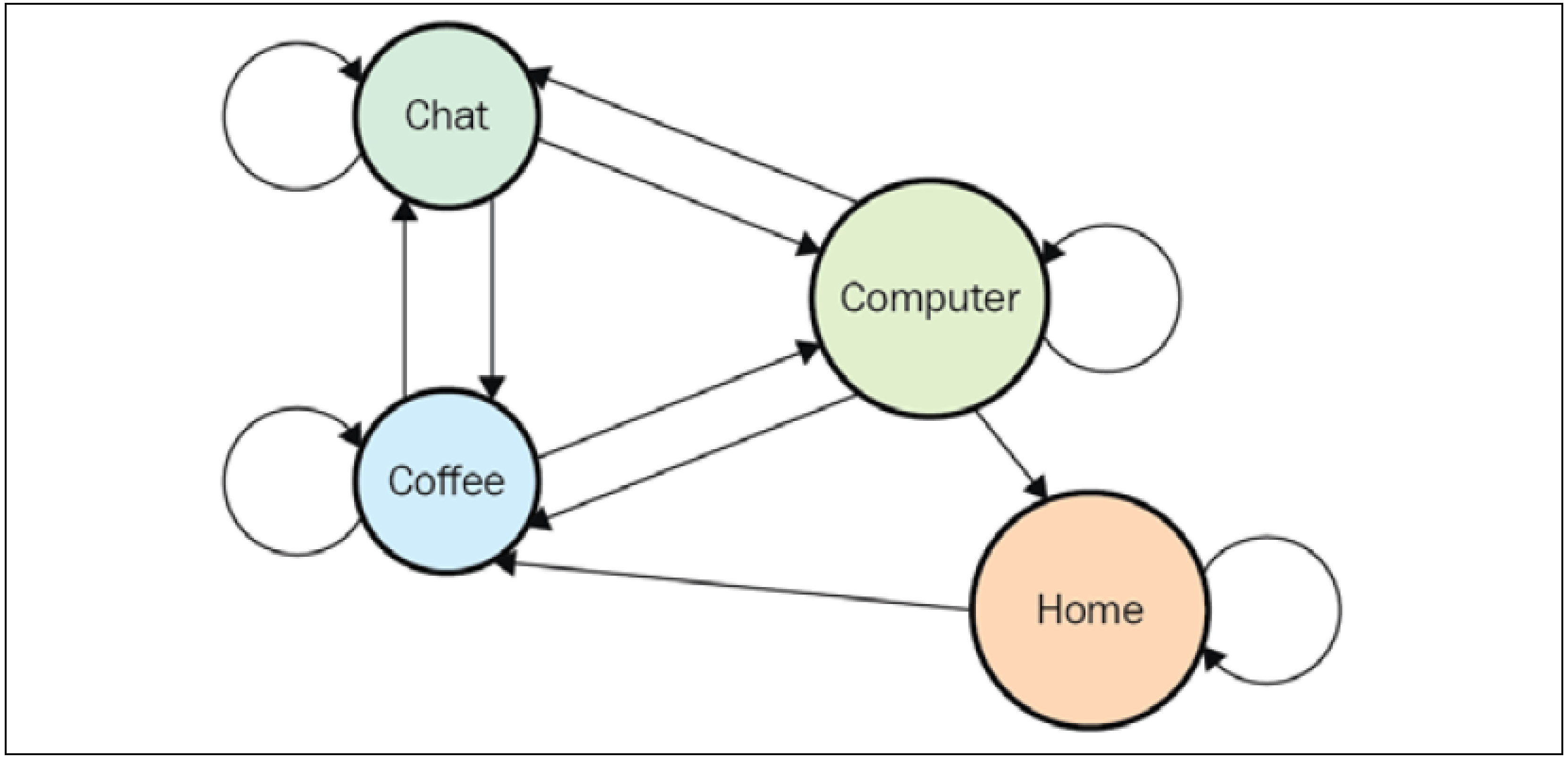
- Set of states S, set of actions A, initial state S_0
- Transition model $P(s,a,s')$
  – P([1,1], up, [1,2]) = 0.8
- Reward function $r(s)$
  – r([4,3]) = +1



- **Goal:** maximize cumulative reward in the long run
- **Policy:** mapping from S to A
  – $\pi(s)$ or $\pi(s,a)$ (deterministic vs. stochastic)
- **Reinforcement learning:**
  – transitions and rewards usually not available
  – how to change the policy based on experience
  – how to explore the environment

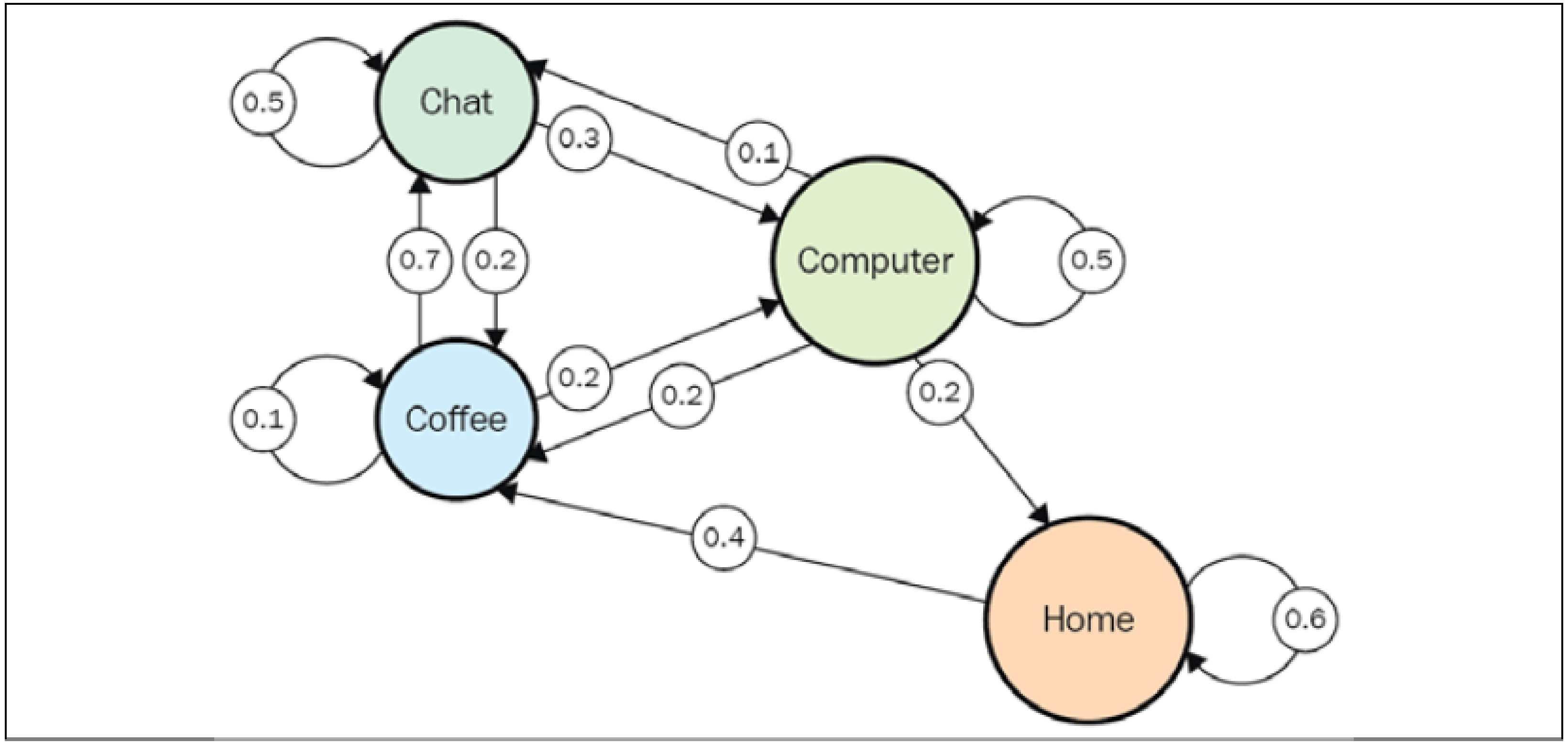
From Reinforcement Learning Tutorial, Peter Bodík, RAD Lab, UC Berkeley

Let’s define states



https://subscription.packtpub.com/book/data/9781838826994/1/ch01lvl1sec07/the-theoretical-foundations-of-rl

And transition probabilities



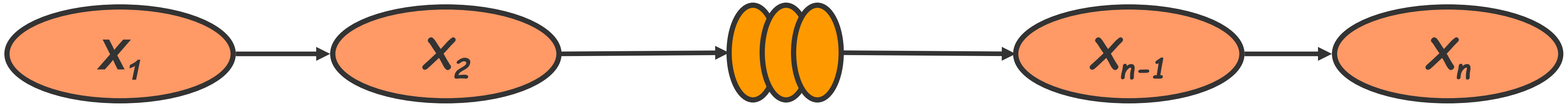
Finite Markov chain

An ***integer time stochastic process***, consisting of a ***domain D*** of $m>1$ states $\{s_1,\ldots,s_m\}$ and

1. An ***m*** dimensional ***initial distribution vector*** $(p(s_1),\ldots, p(s_m))$.
2. An ***m×m transition probabilities matrix M***$= (a_{s_i s_j})$

For example, ***D*** can be the letters $\{A, C, T, G\}$, $p(A)$ the probability of A to be the 1st letter in a sequence, and a_{AG} the probability that G follows A in a sequence.

Markov Chain



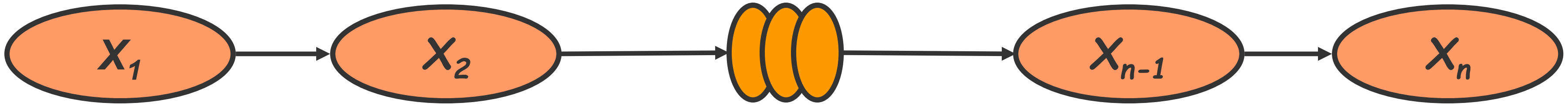
For each integer n, a Markov Chain assigns probability to sequences $(x_1 \ldots x_n)$ over ***D*** (i.e, $x_i \in \boldsymbol{D}$) as follows:

$$p((x_1, x_2, \ldots x_n)) = p(X_1 = x_1) \prod_{i=2}^{n} p(X_i = x_i \mid X_{i-1} = x_{i-1}) = p(x_1) \prod_{i=2}^{n} a_{x_{i-1} x_i}$$

Similarly, $(\boldsymbol{X_1}, \ldots, \boldsymbol{X_i}, \ldots)$ is a sequence of probability distributions over ***D*****.**

There is a rich theory which studies the properties of these sequences.

Markov Chain



Similarly, each $\boldsymbol{X_i}$ is a probability distributions over $\boldsymbol{D}$, which is determined by the initial distribution $(p_1,..,p_n)$ and the transition matrix $\boldsymbol{M}$.

There is a rich theory which studies the properties of such "Markov sequences" $(\boldsymbol{X_1,\ldots, X_i}\ ,\ldots)$.

Matrix Representation

	A	B	C	D
A	0.95	0	0.05	0
B	0.2	0.5	0	0.3
C	0	0.2	0	0.8
D	0	0	1	0

The transition probabilities Matrix $\boldsymbol{M} = (\boldsymbol{a_{st}})$

$\boldsymbol{M}$ is a stochastic Matrix:

$$\sum_t a_{st} = 1$$

The initial ***distribution vector*** $(u_1 \ldots u_m)$ defines the distribution of $\boldsymbol{X_1}$ $(p(\boldsymbol{X_1} = s_i) = u_i)$.

Then after one move, the distribution is changed to $\boldsymbol{X_2} = \boldsymbol{X_1 M}$

Matrix Representation

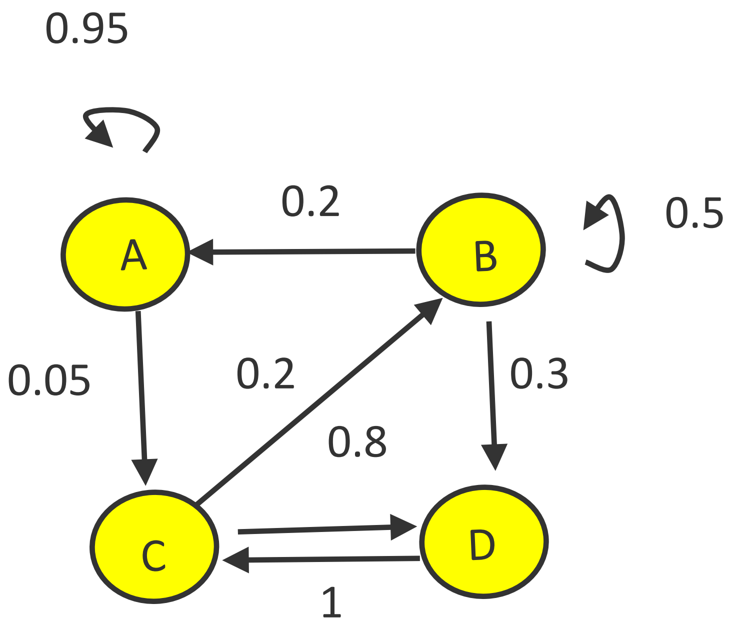
	A	B	C	D
A	0.95	0	0.05	0
B	0.2	0.5	0	0.3
C	0	0.2	0	0.8
D	0	0	1	0

Example: if $\mathbf{X_1}$=(0, 1, 0, 0)
then $\mathbf{X_2}$=(0.2, 0.5, 0, 0.3)

And if $\mathbf{X_1}$=(0, 0, 0.5, 0.5)
then $\mathbf{X_2}$=(0, 0.1, 0.5, 0.4).

The *i*-th distribution is $\boldsymbol{X_i = X_1 M^{i-1}}$

Representation of Markov Chain as Digraph



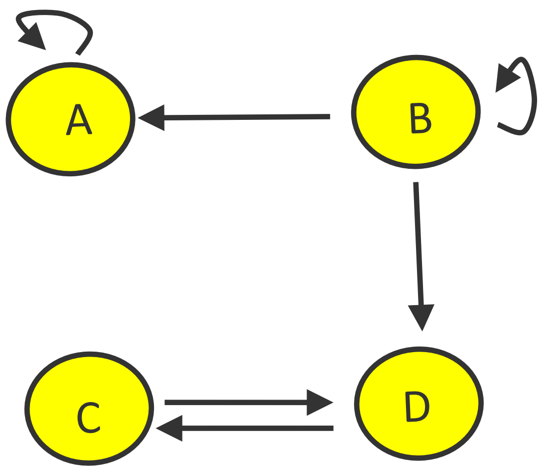
	A	B	C	D
A	0.95	0	0.05	0
B	0.2	0.5	0	0.3
C	0	0.2	0	0.8
D	0	0	1	0

Each directed edge A→B is associated with the **positive** transition probability from A to B.

Properties of Markov Chain States

States of Markov chains are classified by the digraph representation (omitting the actual probability values)

A, C and D are ***recurrent*** states: they are in strongly connected components which are **sinks** in the graph.

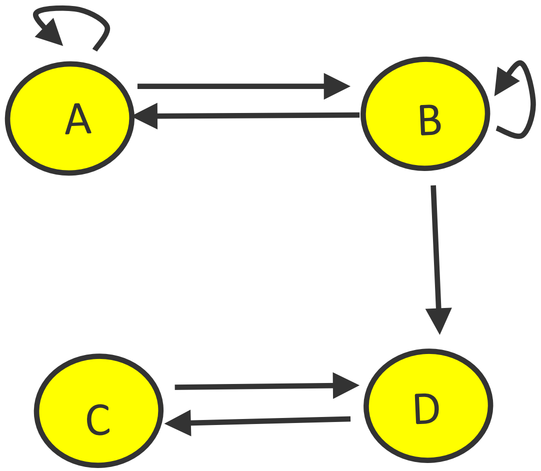


B is not recurrent – it is a ***transient*** state

<u>Alternative definitions:</u>

A state ***s*** is ***recurrent*** if it can be reached from any state reachable from ***s;*** otherwise it is ***transient***.

Another example: transient and recurrent states

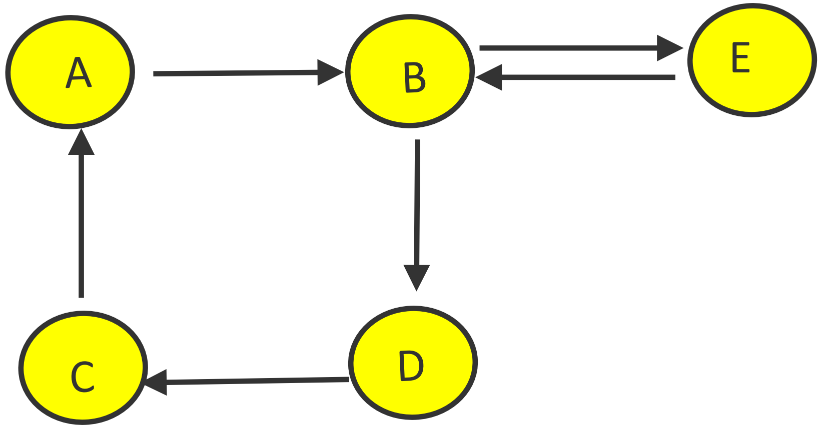


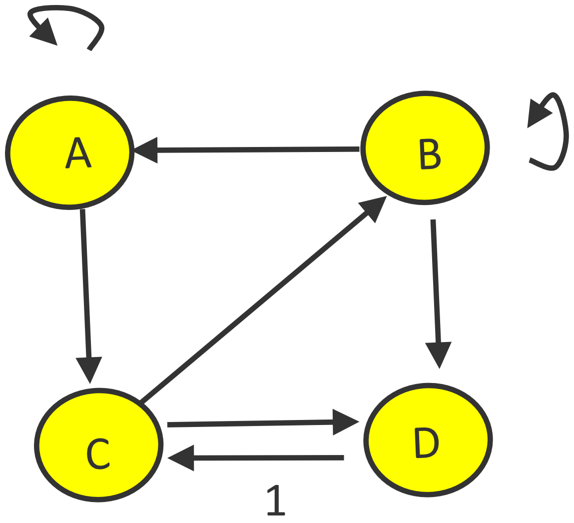
A and **B** are ***transient*** states, **C** and **D** are ***recurrent*** states.

Once the process moves from **B** to **D**, it will never come back.

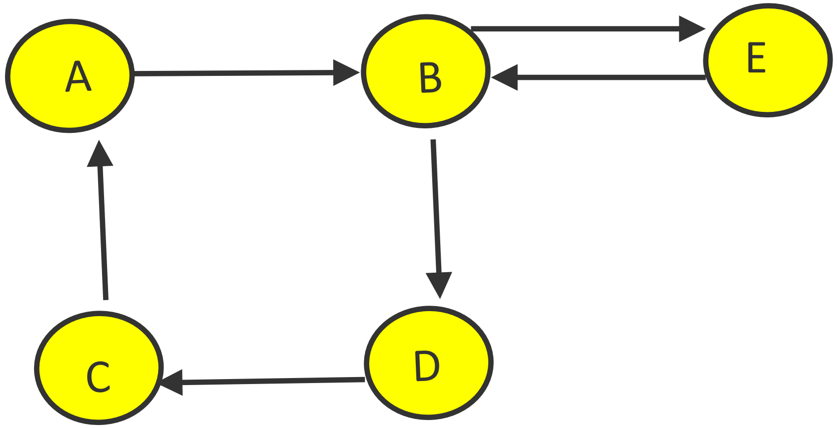
Irreducible Markov Chain

A Markov Chain is ***irreducible*** if the corresponding graph is strongly connected (and thus all its states are recurrent).





Periodic States

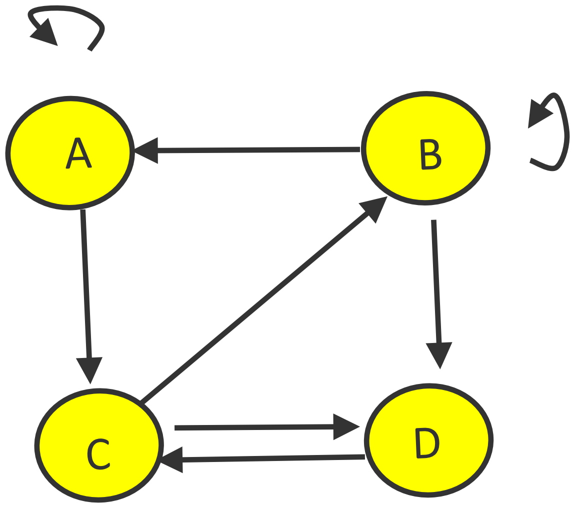


A state *s* has a period ***k*** if ***k*** is the ***GCD*** of the lengths of all the cycles that pass via *s***.** (in the shown graph the period of A is 2).

All the states in the same strongly connected component have the same period

A Markov Chain is ***periodic*** if all the states in it have a period $k > 1$. It is ***aperiodic*** otherwise.

Ergodic Markov Chain



A Markov chain is ***ergodic*** if :

1. ***the corresponding graph is strongly connected.***
2. ***It is not periodic***

Ergodic Markov Chains are important since they guarantee the corresponding Markovian process converges to a unique distribution, in which all states have strictly positive probability.

Stationary Distribution for Markov Chain

Let $\boldsymbol{M}$ be a Markov Chain of m states, and let $\boldsymbol{V} = (v_1,\ldots,v_m)$ be a probability distribution over the m states

$\boldsymbol{V} = (v_1,\ldots,v_m)$ is ***stationary distribution*** for $\boldsymbol{M}$ if $\boldsymbol{VM=V}$.

(i.e., if one step of the process does not change the distribution).

$\boldsymbol{V}$ is a stationary distribution

$\Updownarrow$

$\boldsymbol{V}$ is a left (row) Eigenvector of $\boldsymbol{M}$ with Eigenvalue 1.

Good Markov Chain

A Markov Chains is *good* if the distributions $\boldsymbol{X_i}$, as $\boldsymbol{I} \to \infty$:

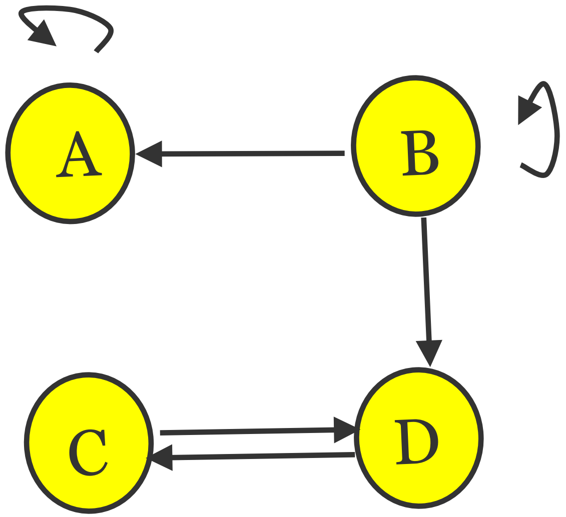
(1) converge to a unique distribution, independent of the initial distribution.

(2) In that unique distribution, each state has a positive probability.

The Fundamental Theorem of Finite Markov Chains:

A Markov Chain is good $\Leftrightarrow$ the corresponding graph is ergodic.

Bad Markov Chains



Consider two initial distributions:

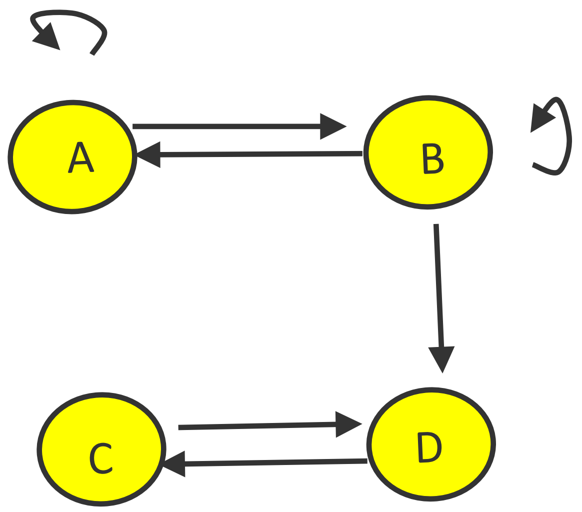
a) $p(\boldsymbol{X}_1 = A) = 1$ ($p(\boldsymbol{X}_1 = \boldsymbol{x}) = 0$ if $\boldsymbol{x} \neq$ A).

b) $p(\boldsymbol{X}_1 =$ C$) = 1$

In case *a)*, the sequence will stay at A forever.
In case *b)*, it will stay in {C,D} for ever.

If G has two states which are unreachable from each other, then $\{\boldsymbol{X}_i\}$ cannot converge to a distribution which is independent on the initial distribution.

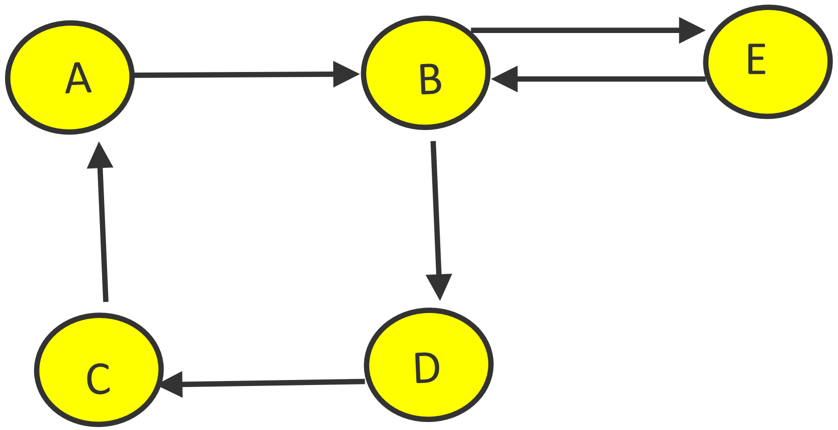
Transient state



Once the process moves from **B** to **D**, it will never come back.

For each initial distribution, with probability 1 a transient state will be visited only a finite number of times.

Periodic Chain



In a periodic Markov Chain (of period k >1) there are initial distributions under which the states are visited in a periodic manner.

Under such initial distributions $\boldsymbol{X_i}$ does not converge as $\boldsymbol{i} \rightarrow \infty$.

Corollary: A good Markov Chain is not periodic

Hidden Markov Model

- A Hidden Markov model is a statistical model in which the system being modelled is assumed to be Markov process with unobserved hidden states.

- In Regular Markov models the state is clearly visible to others in which the state transition probabilities are the parameters only whereas in HMM the state is not visible but the output is visible.

Hidden Markov Model

- It consists of set of states : $S_1, S_2, S_3 \ldots\ldots S_n$.
- Process moves from one state to another state generating a sequence of states Si1, Si2,…. Sik…
- Markov chain property: probability of each subsequent state depends only on what was the previous state

$$P(Sik \mid Sk1, Si2, \ldots\ldots Sik\text{-}1) = P (Sik \mid Sik\text{-}1)$$

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

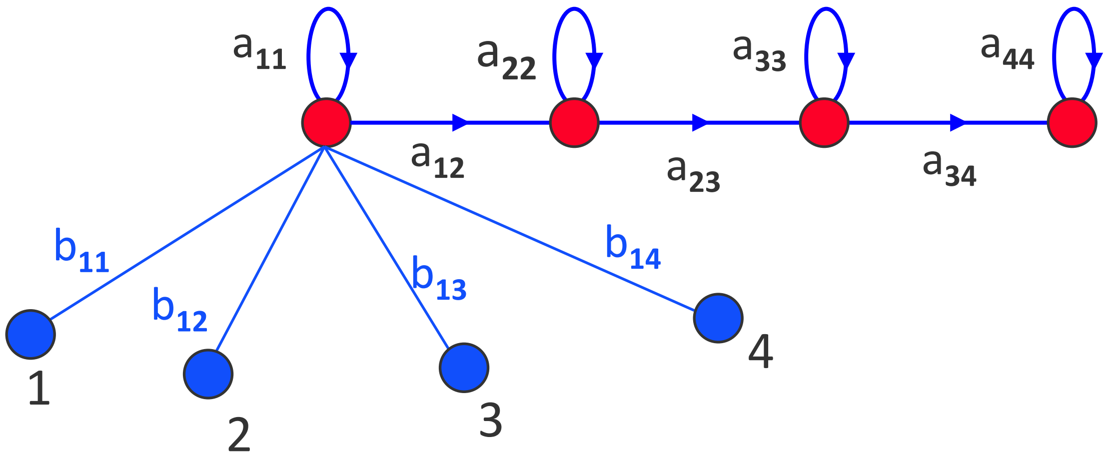
$$V = \{ v1, v2, v3 \ldots\ldots vk \ldots\ldots \}$$

Hidden Markov Model

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

- matrix of transition probabilities $A=(a_{ij})$, $a_{ij}= P(s_i \mid s_j)$,
- matrix of observation probabilities $B=(b_i(v_m))$, $b_i(v_m) = P(v_m \mid s_i)$, and
- a vector of initial probabilities $p=(\pi_i)$, $p_i = P(s_i)$

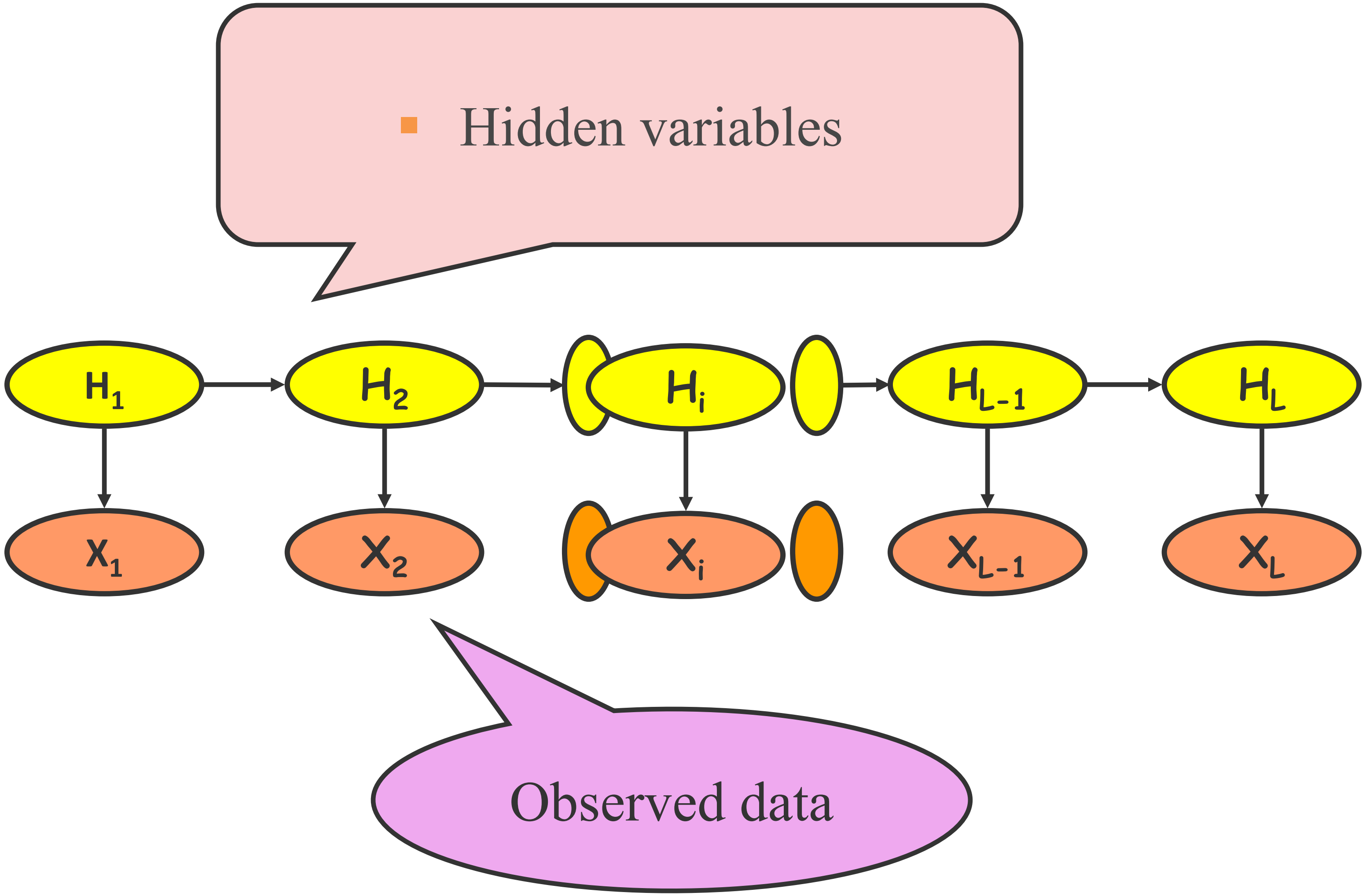
Model is represented by $M=(A, B, p)$.



- a_{ij} are state transition probabilities.
- b_{ik} are observation (output) probabilities.

- $b_{11} + b_{12} + b_{13} + b_{14} = 1$,
- $b_{21} + b_{22} + b_{23} + b_{24} = 1$.

Hidden Markov Model



Main Problems

- **Evaluation problem:** Given the HMM M = { A, B, p } and observation sequence O = o1,o2 ......ok, calculate the probability that model m has generated sequence O.

- **Decoding problem :** Given the HMM M = { A, B, p } and observation sequence O = o1,o2 ......ok, Calculate the most likely sequence of hidden states Si that generated sequence O.

- **Learning Problem :** Given some training observation sequences O = o1,o2 ......ok, and general structure of HMM(visible and hidden states), determine HMM parameters that best fit the training data.