Memory and Machine Learning: course introduction

M1 SCA

Laure Buhry laure.buhry@loria.fr
LORIA bureau C042

Semester 2

Bibliography

Artificial Intelligence : *Artificial Intelligence*, Stuart Russel and Peter Norvig. (pdf version available)
Bayesian Reasoning and Machine Learning David Barber (pdf version available)
http://web4.cs.ucl.ac.uk/staff/D.Barber/pmwiki/pmwiki.php?n=Brml.HomePage

Probabilities and Markov Chains :
Probabilités de l'Ingénieur. Variables aléatoires et simulation, Nicolas Bouleau, Hermann Éditeurs des Sciences et des Arts, 2002.

Bayesian networks :
Réseaux Bayésiens, Patrick Naïm et al., Éditions Eyrolles.
Bayesian networks, Judea Pearl, 1988, Handbook of Brain Theory and Neural Networks, MIT Press, pp. 157–160.

Plan

1. Biological inspiration for inference learning in IA
2. Mathematical tools
   - Conditional probabilities
   - Bayes rule
3. Introduction to bayesian networks
4. Markov chains
   - Definitions and properties
   - Example
   - Properties
5. A “real-world” example : Google
6. Introduction to Hidden Markov Models
7. References

What is bayesian reasoning ?

Deductive / inductive reasoning

Deductive reasoning : reasoning based on a logical proceeding of arguments of the type “cause implies consequence”.

Ex : Socrate is Human, so Socrate is mortal

Inductive reasoning : reasoning using the induction process, i.e. using a sample, an observable variable to formulate a general fact, or making use a the consequence to find the cause

What is bayesian reasoning ?

Deductive / inductive reasoning

Deductive reasoning : reasoning based on a logical proceeding of arguments of the type "cause implies consequence".

Ex : Socrate is Human, so Socrate is mortal

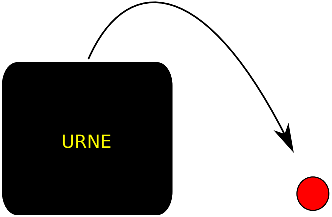
Inductive reasoning : reasoning using the induction process, i.e. using a sample, an observable variable to formulate a general fact, or making use a the consequence to find the cause

Ex : Socrate is mortal, so Socrate is Human
$\rightarrow$ not always true : uncertainty

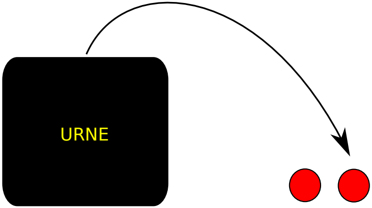
Other exemple of inductive reasoning

contains red and blue balls

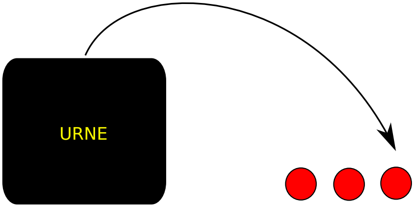
Other exemple of inductive reasoning



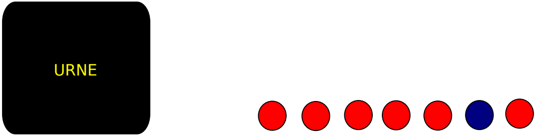
Other exemple of inductive reasoning



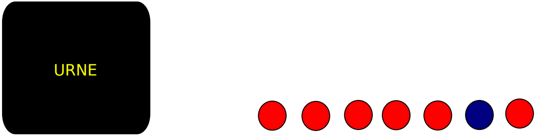
Other exemple of inductive reasoning



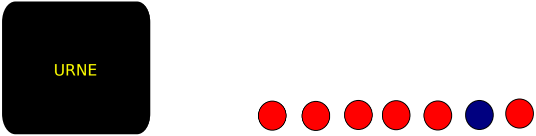
Other exemple of inductive reasoning



Other exemple of inductive reasoning

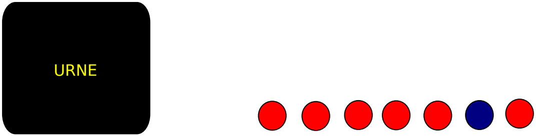


Other exemple of inductive reasoning



Inductive reasoning :

Other exemple of inductive reasoning



Inductive reasoning :
The majority of drawn balls is red, so the box contains more red balls than blue balls.

Mini exercise

Amongst these statements, which ones follow an inductive reasoning ?

a) $x = 20$ and $y = 35$, then $xy = 700$

b) All flowers have petals, iris is a flower, so iris has petals.

c) $x + y = 8$, then $x = 4$ and $y = 4$

d) Plants are living being, my cat is a living being, so my cat is plant.

e) Let A and B two sets so that A is included in B. If a is an element of B, then a is an element of A.

Mini exercise

Amongst these statements, which ones follow an inductive reasoning ?

a) $x = 20$ and $y = 35$, then $xy = 700$

b) All flowers have petals, iris is a flower, so iris has petals.

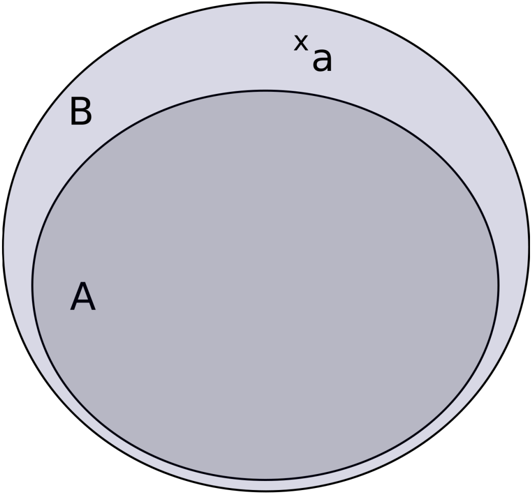
c) $x + y = 8$, then $x = 4$ and $y = 4$

d) Plants are living being, my cat is a living being, so my cat is plant.

e) Let A and B two sets so that A is included in B. If *a* is an element of B, then *a* is an element of A.

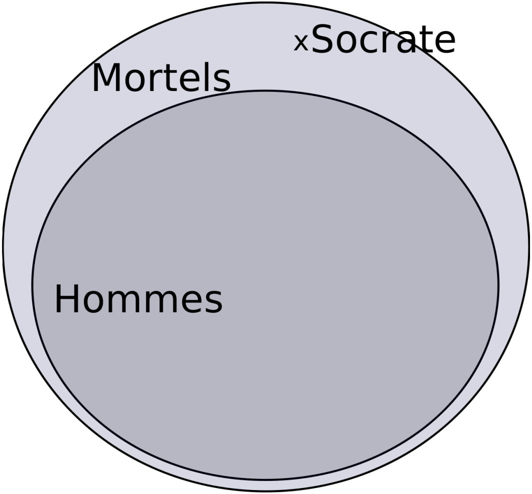
Mini exo

e)



Mini exo

e)



Plan

Plan

Reminder : conditional probabilities

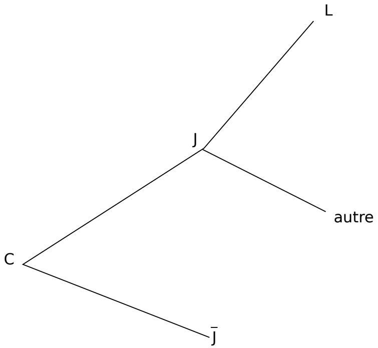
Given two events A and B from the sigma-field of a probability space, the ***conditional probability of A given B*** is defined as the quotient of the probability of the joint of events A and B, and the probability of B :

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

with $\mathbb{P}(B) \neq 0$

Reminder : conditional probabilities

Example : Let us consider the events :



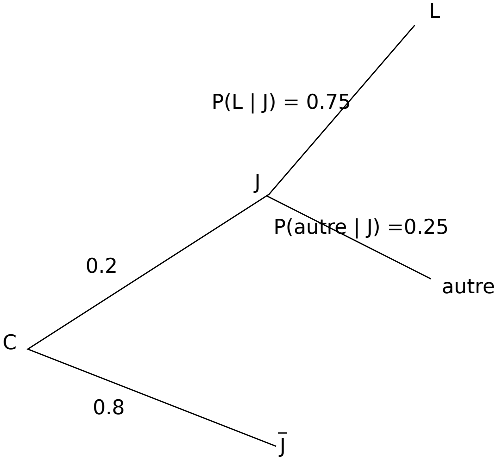
E : "a truck passes"
J : "the truck is yellow"
$\bar{J}$: " the truck is not yellow"
L : "the truck is a truck of the Post Office"

We have here :
$\mathbb{P}(L \cap J) = \mathbb{P}(L) = 0.15$
$\mathbb{P}(J) = 0.2$
$\mathbb{P}(\bar{J}) = 0.8$

What is the value of $\mathbb{P}(L|J)$?

Reminder : conditional probabilities

Exemple :



Reminder : conditional probabilities

Law of total probability

Let $(A_i)_{i \in I}$ aset of events such as $\cup_{i \in I}(B \cap A_i) = B$, $\forall i \neq j$, $A_i \cap A_j = \varnothing$ and $A_i \neq \varnothing$ for all i.

Then :

$$\mathbb{P}(B) = \sum_{i \in I} \mathbb{P}(B \cap A_i) = \sum_{i \in I} \mathbb{P}(B|A_i)\mathbb{P}(A_i)$$

Plan

Reminder : Bayes' rule

Theorem

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

$\mathbb{P}(A)$: ***prior odds*** on A. $\mathbb{P}(A)$ is also called ***marginal probability of A***.

$\mathbb{P}(A|B)$: ***posterior odds*** on A given B (or from A under the condition B) .

$\mathbb{P}(B|A)$, for known B : ***likelihood ratio*** of A.

$\mathbb{P}(B)$: prior odds on B.
$\rightarrow$ The rule tells us how our prior beliefs concerning whether or not A is true needs to be updated on receiving the information B

Reminder : Bayes' rule

Proof

One uses the definition of conditional probability and one expresses the probability of the intersection between two events A and B ; there are two ways of writing it :

$$\mathbb{P}(A \cap B) = \mathbb{P}(\mathbf{A}|\mathbf{B})\mathbb{P}(\mathbf{B})$$

or

$$\mathbb{P}(A \cap B) = \mathbb{P}(\mathbf{B}|\mathbf{A})\mathbb{P}(\mathbf{A})$$

Hence :

$$\mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A)$$

Therefore, if the probability of B is non-zero :

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

Interpretation of the Bayes'rule

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

One can see :
$\mathbb{P}(A)$ as the degree of confidence of the hypothesis A before taking the observations into account

Interpretation of the Bayes'rule

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

One can see :
$\mathbb{P}(A)$ as the degree of confidence of the hypothesis A before taking the observations into account

$\mathbb{P}(A|B)$ as the degree of confidence once we have taken into account the observations

Interpretation of the Bayes'rule

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

One can see :
$\mathbb{P}(A)$ as the degree of confidence of the hypothesis A before taking the observations into account

$\mathbb{P}(A|B)$ as the degree of confidence once we have taken into account the observations

$\mathbb{P}(B|A)$: quantifies the degree of compatibility between hypothesis A and observation B

Interpretation of the Bayes'rule

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

One can see :
$\mathbb{P}(A)$ as the degree of confidence of the hypothesis A before taking the observations into account

$\mathbb{P}(A|B)$ as the degree of confidence once we have taken into account the observations

$\mathbb{P}(B|A)$: quantifies the degree of compatibility between hypothesis A and observation B

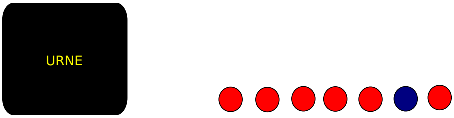
$\rightarrow$ allows to revise ones confidence according to the observation and to quantify uncertainty

Back to the box example

contiains red and blue balls

$\rightarrow$ hypothesis A " drawing a red ball" has prior odds equal to "drawing a blue ball" : $\mathbb{P}(A) = 0.5$

Back to the box example



$\rightarrow$ posterior hypothesis $A|B$ "drawing a red ball given the observation (6 red balls were drawn over the 7 balls)"

Bayes' rule and inductive reasoning ?

Reasoning based on Bayes'rule is inductive reasoning.

Indeed, inductive reasoning consists in observing ("More red balls than blue balls were drawn") and in considering that the observations support an hypothesis ("There is more red balls than blue balls in the box").

Exercise

Let us consider an infection screening. If a patient contracts the tested disease, the screening is positive in 99% f the cases. If a patient is healthy, the screening is correct, i.e. negative in 95% of the cases.

Imagine that the disease only affects one over one thousand personne, i.e. with a probability of 0.001.

What is the probability that the screening provides a false positive ?

Exercise – Correction

Formalisation :

Let A the event "The patient contracted the disease"and B the event "The test is positive".

One looks for $\mathbb{P}(\bar{A}|B)$ (where $\bar{A}$ is the complementary set of A, i.e. non-sick = healthy)

Exercise – Correction

Formalisation :
Let A the event "The patient contracted the disease"and B the event "The test is positive".

One looks for $\mathbb{P}(\bar{A}|B)$ (where $\bar{A}$ is the complementary set of A, i.e. non-sick = healthy)

But, $\mathbb{P}(\bar{A}|B) = 1 - \mathbb{P}(A|B)$

Exercise – Correction

Formalisation :

Let A the event "The patient contracted the disease"and B the event "The test is positive".

One looks for $\mathbb{P}(\bar{A}|B)$ (where $\bar{A}$ is the complementary set of A, i.e. non-sick $=$ healthy)

But, $\mathbb{P}(\bar{A}|B) \; = \; 1 - \mathbb{P}(A|B)$

and the wording gives : $\mathbb{P}(B|A) \; = \; 0.99$
$\mathbb{P}(A) \; = \; 0.001$
$\mathbb{P}(\bar{B}|\bar{A}) \; = \; 0.95$

Exercise – Correction

Formalisation :

Let A the event "The patient contracted the disease"and B the event "The test is positive".

One looks for $\mathbb{P}(\bar{A}|B)$ (where $\bar{A}$ is the complementary set of A, i.e. non-sick = healthy)

But, $\mathbb{P}(\bar{A}|B) = 1 - \mathbb{P}(A|B)$

and the wording gives : $\mathbb{P}(B|A) = 0.99$
$\mathbb{P}(A) = 0.001$
$\mathbb{P}(\bar{B}|\bar{A}) = 0.95$

Let us then calculate $\mathbb{P}(A|B)$

Exercise – Correction

Moreover, thanks to Bayes'rule :

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)} \quad (1)$$

Also one must first determine $\mathbb{P}(B)$.

Exercise – Correction

Moreover, thanks to Bayes'rule :

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)} \tag{1}$$

Also one must first determine $\mathbb{P}(B)$.

In absolute terms, the screening can be positive in two cases : if the patient is healthy or if he is sick. In other words, $B = (B \setminus A) \cup (B \setminus \bar{A})$.

And then :$\mathbb{P}(B) = \mathbb{P}(B|A) \times \mathbb{P}(A) + \mathbb{P}(B|\bar{A}) \times \mathbb{P}(\bar{A})$ (total probability)

Exercise – Correction

Moreover, thanks to Bayes'rule :

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)} \quad (1)$$

Also one must first determine $\mathbb{P}(B)$.

In absolute terms, the screening can be positive in two cases : if the patient is healthy or if he is sick. In other words, $B = (B \setminus A) \cup (B \setminus \bar{A})$.

And then :$\mathbb{P}(B) = \mathbb{P}(B|A) \times \mathbb{P}(A) + \mathbb{P}(B|\bar{A}) \times \mathbb{P}(\bar{A})$ (total probability)

As $\mathbb{P}(B|\bar{A}) = 1 - \mathbb{P}(\bar{B}|\bar{A})$,
one has : $\mathbb{P}(B) = 0.001 \times 0.99 + (1 - 0.001) \times 0.05$

Exercise – Correction

$$\mathbb{P}(A|B) = \frac{0.99 \times 0.001}{0.99 \times 0.001 + 0.05 \times 0.999},$$
$$\simeq 0.019$$

Therefore, given that the test is positive, the probability that the patient is healthy is $1 - \mathbb{P}(A|B) \simeq 1 - 0.019 = 0.981$.

Base rate fallacy

Base rate fallacy or **base rate neglect** or **base rate bias**

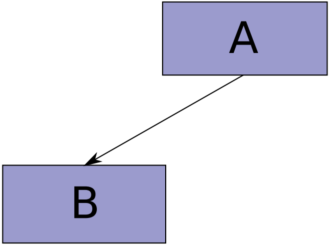
The result of the previous exercise illustrates this well-known bias

In this example, we forget that the probability of having the disease is very low

Plan

Bayesian network : definition

Bayesian network is a graphical modele representing the probability relations (uncertainty) between events. It generally allows to represent causal links between two events.



If there exist a causal relation from A to B, then all information on A can change our knowledge of B and reciprocally.

Bayesian network : formal definition

Definition : a bayesian network is defined by :

- an oriented acyclic graph G, $C = G(V, E)$ where V is the set of nodes and E the set of edges of the graph.
- a probability space $(\Omega, \mathscr{A}, p)$
- a set of n random variables on (Ω, p) associated with the nodes of G such that :
  $p(V) = p(V_1 \cup V_2 \cup \ldots \cup V_n) = \prod_{i=1\ldots n} p(V_i | parents(V_i))$

Bayesian network : use

- problem modelling

  $\rightarrow$ gives a view, a global representation of the problem to be solved

- learning and prediction (inference)

  $\rightarrow$ analysis of relations between events for decision making for instance : largely used by expert systems

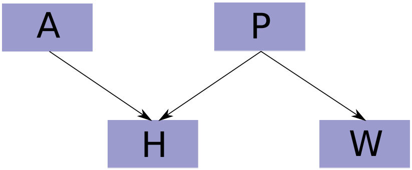
Example

Mr. Holmes notices that his garden's grass is wet : event denoted as H

There can be to reasons :
either the sprinkler was not turned off, event S,
or it rained, event R.

He then looks at his neighbors garden, M. Watson : the grass is wet too, event W

Translation into a causal graph :

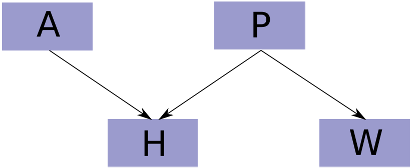


Extract from (Pearl 1988)

Example

S	
T	F
0,4	0,6

R	
T	F
0,4	0,6



	S = T		S = F	
	R = T	R = F	R= T	R = F
H = T	1	1	1	0
H = F	0	0	0	1

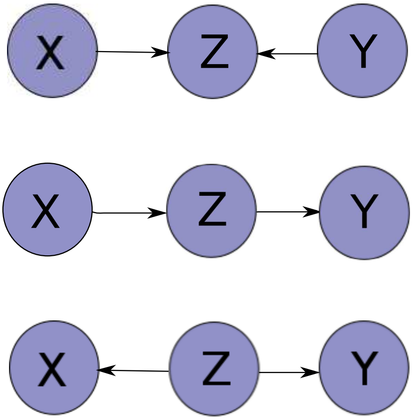
	R = T	R = F
W = T	1	0
W = F	0	1

Question asked by M. Holmes : $\mathbb{P}(A|H)$

Solving : computation thanks to the Bayes' formula

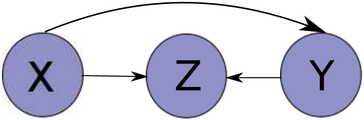
Information flow

Let 3 events X, Y, Z, one can have the following causalgraphs associated with bayesian networks :

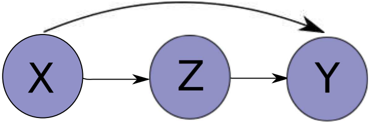


Information flow

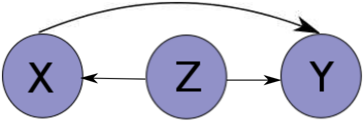
Let 3 events X, Y, Z, one can have the following causalgraphs associated with bayesian networks :



The information can only flow from X to Y if Z is known



The information can only flow from X to Y if Z is not known



The information can only flow from X to Y if Z is not known

D-sÃ©paration

Two events X et Y are said d-separated by Z if, for all trail P between X and Y, a) or b) holds :

a) the trail converges in a node W, such that $W \neq Z$ and W is not a direct cause of Z

b) the trail passes by Z and is, either divergent, or in a chain with Z.

Bayesian network : advantages and drawbacks

- Drawbacks
  - One always has to do a lot of calculations with Bayes's rule
  - Difficulty in defining proability tables associated with the graph
- Advantages
  - The graph representation gives a global view of the problem to be solved ;
  - It allows the use of graph computation tools and graph search (powerful tools)

Plan

Plan

Markov chains : definition

Let E be a countable or measurable set, $(\Omega, \mathscr{A}, \mathbb{P})$ a probability space and $X_0, X_1, \cdots, X_n$ a suite of E-valued random defined on $(\Omega, \mathscr{A}, \mathbb{P})$.

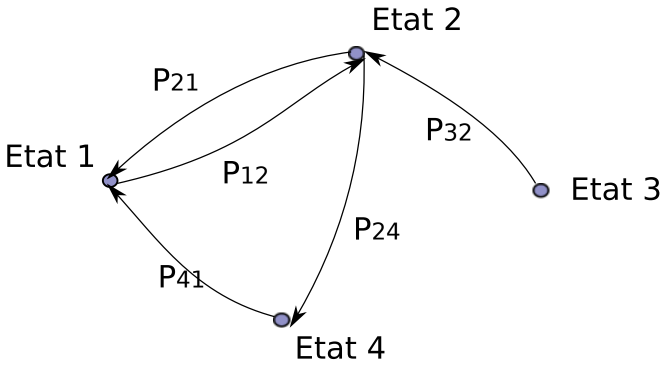
$(X_0, X_1, \cdots, X_n)$ is a homogeneous ***Markov chain*** iff, for all n in $\mathbb{N}$, for all i_0, i_1, ..., i_{n-1}, i, j in E, we have :
$P_{ij} = \mathbb{P}(X_{n+1} = j | X_0 = i_0, X_1 = i_1, \cdots, X_{n-1} = i_{n-1}, X_n = i) = \mathbb{P}(X_{n+1} = j | X_n = i)$

and if P_{ij} does not depend on n.

$\rightarrow$ Usually used to represent events that evolve in time.

$\rightarrow$ the value at time $n+1$ does only depend on the value at time n.

Diagram representation



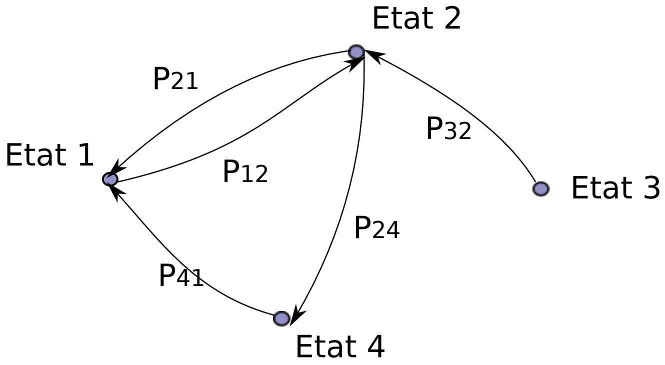
Transition probabilities

The P_{ij} are called ***transition probabilities***.

They represent the probability of changing from one state i to another state j.

The transition are often summarized in a ***transition matrix*** $P = (P_{ij})$.

Diagram representation and transition matrix



$$\begin{pmatrix} 0 & P_{12} & 0 & 0 \\ P_{21} & 0 & 0 & P_{24} \\ 0 & P_{32} & 0 & 0 \\ P_{41} & 0 & 0 & 0 \end{pmatrix}$$

Proposition

In order to ease the calculations on transition matrices, one uses the following proposition :

Proposition :

a) $\mathbb{P}(X_n = j | X_0 = i) = (P^n)_{ij}$

b) $\mathbb{P}(X_n = j) = \sum_i \mathbb{P}(X_0 = i)(P^n)_{ij}$

Proof

Proof :

i) One assumes that space E is finite, then P is a square matrix.

One demonstrates a) by mathematical induction (récurrence) : one assumes that a) holds for $n = 1$.

Let us assume it also holds for n and calculate $\mathbb{P}(X_{n+1} = j | X_0 = i)$:

$$\mathbb{P}(X_{n+1} = j | X_0 = i) = \sum_k \mathbb{P}(X_{n+1} = j, X_n = k | X_0 = i)$$

Thus, (using the definition of conditional probabilities) :

$$\mathbb{P}(X_{n+1} = j | X_0 = i) = \sum_k \mathbb{P}(X_{n+1} = j | X_n = k, X_0 = i)\mathbb{P}(X_n = k | X_0 = i)$$

Proof

As (X_l) is a Markov chain :

$$\mathbb{P}(X_{n+1} = j | X_0 = i) = \sum_k \mathbb{P}(X_{n+1} = j | X_n = k) \mathbb{P}(X_n = k | X_0 = i)$$

Therefore, according to the induction hypothesis :

$$\mathbb{P}(X_{n+1} = j | X_0 = i) = \sum_k \mathbb{P}(X_{n+1} = j | X_n = k)(P^n)_{ik} = \sum_k P_{kj}(P^n)_{ik}$$

One recognize the formula of a matrix product, so :

$$\mathbb{P}(X_{n+1} = j | X_0 = i) = \sum_k \mathbb{P}(X_{n+1} = j | X_n = k)(P^n)_{ik} = (P^{n+1})_{ij}$$

Proof

ii) When E is countable, P is a stochastic matrix with finite dimension, in other words, this is a map from $E \times E$ to [0,1] :
$i, j \to P_{ij}$ so that
$\forall i \in E, \sum_j P_{ij} = 1$

Infinite stochastic matrices multiply as ordinary matrices, also we can define, for two infinite stochastic matrices P and Q :

$$(PQ)_{ij} = \sum_{k \in E} P_{ik} Q_{kj}$$

The right hand side is always convergent because $Q_{kj} \leq 1$, so :

$$\sum_j [\sum_{k \in E} P_{ik} Q_{kj}] = \sum_k [\sum_j P_{ik} Q_{kj}] = \sum_k P_{ik} = 1$$

Hence we have a stochastic matrix PQ and one can apply the reasoning used in i).

Plan

Example

Agent : Leo cat*

Leo can essentially be in one of the three following *states* :
bed on the couch (C), eating cat nuggets (M), walk in the garden (J).

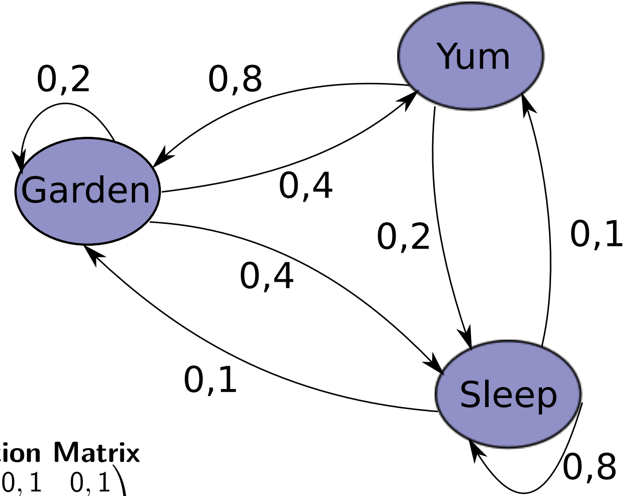
*any resemblance to actual persons living or dead is purely coincidental

Example

Transition probabilities from one state to another are :

- When Leo is sleeping, there are 8 chances out of 10 that it does not wake up in the following minute
- When it wakes up, there is one chance out of 2 that it eats and one chance out of 2 that it goes to the garden
- When it ate, there are 2 chances out of 10 that it goes back to sleep and 8 chances out of 10 that it goes to the garden
- Once in the garden, there are 2 chances out of 10 that it stays outside, 4 chances out of 10 that he goes back to sleep, and 4 chances out of 10 that he eats

Example - Formalisation



Transition Matrix

$$\begin{pmatrix} 0,8 & 0,1 & 0,1 \\ 0,2 & 0 & 0,8 \\ 0,4 & 0,4 & 0,2 \end{pmatrix}$$

1 : sleep, 2 : yum, 3 : garden

Example - Prediction

Let us denote x^i Leo's state at time i

Let us suppose that Leo is sleeping during the first minute of observation.

In that case : $x^{(0)} = [1\,0\,0]$

One predicts thus that after one minute :

Example - Prediction

Let us denote x^i Leo's state at time i

Let us suppose that Leo is sleeping during the first minute of observation.

In that case : $x^{(0)} = [1\,0\,0]$

One predicts thus that after one minute :

$$x^{(1)} = x^{(0)}P = [1\,0\,0]\begin{pmatrix} 0,8 & 0,1 & 0,1 \\ 0,2 & 0 & 0,8 \\ 0,4 & 0,4 & 0,2 \end{pmatrix} = [0,8\,0,1\,0,1]$$

Example - Prediction

Let us denote x^i Leo's state at time i

Let us suppose that Leo is sleeping during the first minute of observation.

In that case : $x^{(0)} = [1\,0\,0]$

One predicts thus that after one minute :

$$x^{(1)} = x^{(0)}P = [1\,0\,0]\begin{pmatrix} 0,8 & 0,1 & 0,1 \\ 0,2 & 0 & 0,8 \\ 0,4 & 0,4 & 0,2 \end{pmatrix} = [0,8\,0,1\,0,1]$$

In other words, at time 1, after 1 minute, there are 8 chances out of 10 that it is still sleeping, 1 chance out of 10 that it is eating, and 1 chance out of 10 that it is going to the garden.

Example - Prediction

After two minutes :

$$x^{(2)} = x^{(1)}P = [1\,0\,0]\begin{pmatrix} 0,8 & 0,1 & 0,1 \\ 0,2 & 0 & 0,8 \\ 0,4 & 0,4 & 0,2 \end{pmatrix}^2 = \cdots$$

Plan

Irreducible Markov Chains

Definitions :

A state j is said to be ***accessible*** from a state i iff there exists a $n \geq 0$ such that $P_{ij}^{(n)} > 0$

Irreducible Markov Chains

Definitions :

A state j is said to be ***accessible*** from a state i iff there exists a $n \geq 0$ such that $P_{ij}^{(n)} > 0$

If all states of a Markov chain are accessible, i.e. if it is possible to get to any state from any state, then the chain is said to be ***irreducible*** and all states are said to belong to a same communicating class.

Recurrent state, absorbing state

A state is said to be ***recurrent*** or ***persistent*** iff the probability of getting back to this state is equal to 1. Otherwise, the state is said to be ***transient***.

Recurrent state, absorbing state

A state is said to be ***recurrent*** or ***persistent*** iff the probability of getting back to this state is equal to 1. Otherwise, the state is said to be ***transient***.

A state is said ***absorbing*** state iff the probability to stay in this state is equal to 1.

Remarks

Consequently, a recurrent state can be visited an infinite number of times

Remarks

Consequently, a recurrent state can be visited an infinite number of times

A transient state is only visited a finite number of times

Remarks

Consequently, a recurrent state can be visited an infinite number of times

A transient state is only visited a finite number of times

If a state is recurrent, then all states in its class are recurrent too

Remarks

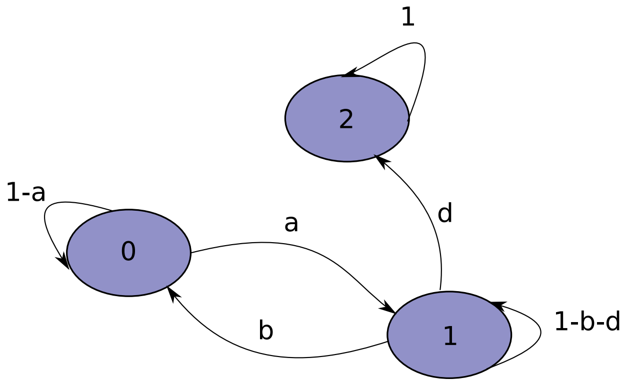
Consequently, a recurrent state can be visited an infinite number of times

A transient state is only visited a finite number of times

If a state is recurrent, then all states in its class are recurrent too

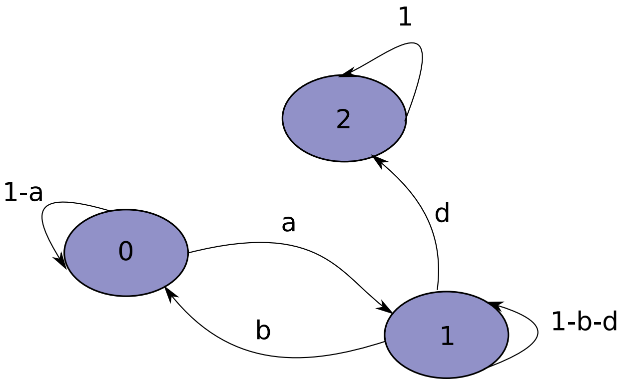
If the initial state of a Markov chain is recurrent, one will visit it an infinite number of times

Example 1



$a, b, d \in [0, 1]$

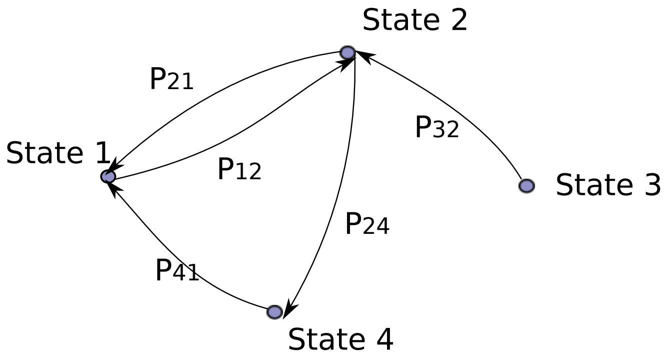
Example 1



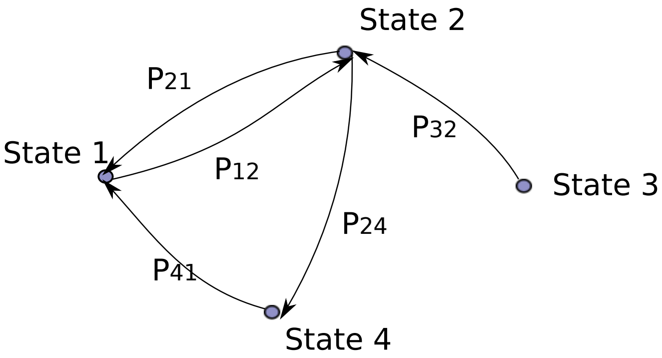
$a, b, d \in [0, 1]$

Here, state 2 is absorbing, but 0 and 1 are transient.

Example 2



Example 2



States 1, 2 and 4 are recurrent. State 3 is transient.

Some More Definitions

Markov Chains of Order k

Let $k \in \mathbb{N}$ be a finite number. A ***Markov chain of order*** k (or ***with memory*** k) is a stochastic process satisfying :

$$\mathbb{P}(X_n = i_n | X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, \ldots, X_1 = i_1) =$$
$$\mathbb{P}(X_n = i_n | X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, \ldots, X_{n-k} = i_{n-k}) \text{ for } n > k$$

$\rightarrow$ the future state depends on the past k states.

It is possible to construct a chain (Y_n) from (X_n) which has the "classical" Markov property by taking as state space the ordered k-tuples of X values, i.e. $Y_n = (X_n, X_{n-1}, \ldots, X_{n-k+1})$.

Going further

Markov chains can be extended to the continuous-time case. A mathematical model which takes values in some finite or countable (measurable) set and for which the time spent in each state takes non-negative real values and has an exponential distribution is called a ***continuous-time Markov chain***.

As for the discrete-time Markov chains, a continuous-time Markov chain is a continuous-time stochastic process with the Markov property which means that future behaviour of the model depends only on the current state of the model and not on historical behaviour.

Use of Markov models

- Learning to predict (economy, weather forecast, etc)
- Modelling queue processes (printers, malls, etc)
- Search engines
- ...

Plan

A “real-world” application : Google [1]

1. http ://blog.kleinproject.org/ ?p=280

Google search : Markov chains and eigenvalues

Contexte : quantity of pages on the World-Wide-Web

$\rightarrow$ thousands to millions of search results

$\rightarrow$ search results have to be ordered to be exploited

$\Rightarrow$ need for a page rank

Google search : Markov chains and eigenvalues

Example search : Neurosys Loria

$\rightarrow$ the front page comes first :
http ://neurosys.loria.fr/Main/HomePage

but one can expect, if people are generally more interested in Neurosys' research projects, that after a few month, the url :
http ://neurosys.loria.fr/Projects/Projects
will come first

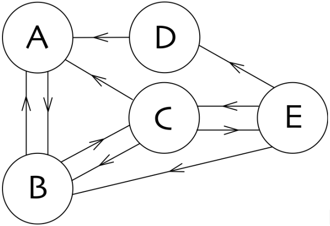
How does that work ?

Google search : Markov chains and eigenvalues

Model

web = oriented graph with vertices that are the pages and oriented edges that are the links between pages

Example



Klein Project

Google search : Markov chains and eigenvalues

One starts a random walk in the graph. We have the following transition matrix P :

$$P = \begin{array}{cc} & \begin{array}{ccccc} A & B & C & D & E \end{array} \\ & \left(\begin{array}{ccccc} 0 & \frac{1}{2} & \frac{1}{3} & 1 & 0 \\ 1 & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & \frac{1}{3} & 0 & 0 \end{array} \right) \begin{array}{c} A \\ B \\ C \\ D \\ E \end{array} \end{array}$$

Google search : Markov chains and eigenvalues

One starts a random walk in the graph. We have the following transition matrix P :

$$P = \begin{array}{cc} & \begin{array}{ccccc} A & B & C & D & E \end{array} \\ & \left(\begin{array}{ccccc} 0 & \frac{1}{2} & \frac{1}{3} & 1 & 0 \\ 1 & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & \frac{1}{3} & 0 & 0 \end{array} \right) \begin{array}{c} A \\ B \\ C \\ D \\ E \end{array} \end{array}$$

Let us consider a random variable X_n with values in the set of pages $\{A, B, C, D, E\}$ ($N = 5$ pages). X_n represents the page where we are after n steps of the random walk.

Google search : Markov chains and eigenvalues

One starts a random walk in the graph. We have the following transition matrix P :

$$
\begin{array}{cc}
 & \begin{array}{ccccc} A & B & C & D & E \end{array} \\
P = & \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{3} & 1 & 0 \\ 1 & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & \frac{1}{3} & 0 & 0 \end{pmatrix} \begin{array}{c} A \\ B \\ C \\ D \\ E \end{array}
\end{array}
$$

Let us consider a random variable X_n with values in the set of pages $\{A, B, C, D, E\}$ ($N = 5$ pages). X_n represents the page where we are after n steps of the random walk.
Hence, if $P = (p_{ij})_{i,j=[1,2,\dots,5]}$, then p_{ij} is the conditional probability that we are on the i-th page at step $n+1$, given that we were on the j-th page at step n : $p_{ij} = \mathbb{P}(X_{n+1} = i \mid X_n = j)$

Google search : Markov chains and eigenvalues

One starts a random walk in the graph. We have the following transition matrix P :

$$P = \begin{array}{cc} & \begin{array}{ccccc} A & B & C & D & E \end{array} \\ & \left(\begin{array}{ccccc} 0 & \frac{1}{2} & \frac{1}{3} & 1 & 0 \\ 1 & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & \frac{1}{3} & 0 & 0 \end{array} \right) \begin{array}{c} A \\ B \\ C \\ D \\ E \end{array} \end{array}$$

Let us consider a random variable X_n with values in the set of pages $\{A, B, C, D, E\}$ ($N = 5$ pages). X_n represents the page where we are after n steps of the random walk.
Hence, if $P = (p_{ij})_{i,j=[1,2,\ldots,5]}$, then p_{ij} is the conditional probability that we are on the i-th page at step $n+1$, given that we were on the j-th page at step n : $p_{ij} = \mathbb{P}(X_{n+1} = i \mid X_n = j)$

$\rightarrow$ (X_n) **Markov chain !**

Google search : Markov chains and eigenvalues

At time 2, one gets :

$$P^2 = \begin{array}{c} \\ \\ \end{array}\begin{array}{cc} \begin{array}{ccccc} A & B & C & D & E \end{array} & \\ \begin{pmatrix} \frac{1}{2} & \frac{1}{6} & \frac{1}{6} & 0 & \frac{11}{18} \\ 0 & \frac{2}{3} & \frac{4}{9} & 1 & \frac{1}{9} \\ \frac{1}{2} & 0 & \frac{5}{18} & 0 & \frac{1}{6} \\ 0 & 0 & \frac{1}{9} & 0 & 0 \\ 0 & \frac{1}{6} & 0 & 0 & \frac{1}{9} \end{pmatrix} & \begin{array}{c} A \\ B \\ C \\ D \\ E \end{array} \end{array}$$

Google search : Markov chains and eigenvalues

At time 2, one gets :

$$P^2 = \begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{pmatrix} \frac{1}{2} & \frac{1}{6} & \frac{1}{6} & 0 & \frac{11}{18} \\ 0 & \frac{2}{3} & \frac{4}{9} & 1 & \frac{1}{9} \\ \frac{1}{2} & 0 & \frac{5}{18} & 0 & \frac{1}{6} \\ 0 & 0 & \frac{1}{9} & 0 & 0 \\ 0 & \frac{1}{6} & 0 & 0 & \frac{1}{9} \end{pmatrix} \end{array} \begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}$$

After about 35 iterations, one finds :

$$P^{35} = \begin{array}{c} \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{pmatrix} 0.293 & 0.293 & 0.293 & 0.293 & 0.293 \\ 0.390 & 0.390 & 0.390 & 0.390 & 0.390 \\ 0.220 & 0.220 & 0.220 & 0.220 & 0.220 \\ 0.024 & 0.024 & 0.024 & 0.024 & 0.024 \\ 0.073 & 0.073 & 0.073 & 0.073 & 0.073 \end{pmatrix} \end{array} \begin{array}{c} \\ A \\ B \\ C \\ D \\ E \end{array}$$

whatever the initial state

Google search : Markov chains and eigenvalues

Let us consider the vector $\pi^t = (0.293, 0.390, 0.220, 0.024, 0.073)$
so that : $P\pi = \pi$ (stationary distribution)

Google search : Markov chains and eigenvalues

Let us consider the vector $\pi^t = (0.293, 0.390, 0.220, 0.024, 0.073)$ so that : $P\pi = \pi$ (stationary distribution)

One can then order the pages according to the highest probability of being in one state after n steps :
B, A, C, E, D

Google search : Markov chains and eigenvalues

General case

web = oriented graph in which the N vertices represent the N pages of the web, and the oriented edges represent the links between pages. The transition probabilities of this graph are summarized in an $N \times N$ matrix, P.

Google search : Markov chains and eigenvalues

General case

web = oriented graph in which the N vertices represent the N pages of the web, and the oriented edges represent the links between pages. The transition probabilities of this graph are summarized in an $N \times N$ matrix, P.

In our example we found a vector π satisfying $P\pi = \pi$.
$\rightarrow$ an eigenvector of the eigenvalue 1.

Google search : Markov chains and eigenvalues

In case you would have forgotten...

Eigenvalue* - *Eigenvector

Let P be an $N \times N$ matrix. A scalar $\lambda \in \mathbb{C}$ is an ***eigenvalue of P*** if there exists a nonzero vector $X \in \mathbb{C}^N$ such that $PX = \lambda X$. Any such vector X is called an ***eigenvector*** of P.

Google search : Markov chains and eigenvalues

In case you would have forgotten...

Eigenvalue - Eigenvector

Let P be an$N \times N$ matrix. A scalar $\lambda \in \mathbb{C}$ is an ***eigenvalue of P*** if there exists a nonzero vector $X \in \mathbb{C}^N$ such that $PX = \lambda X$. Any such vector X is called an ***eigenvector*** of P.

Proposition

Let P be an $N \times N$ matrix. The eigenvalues of P are the roots of the characteristic polynomial $\det(\lambda I - P) = 0$, where I is the $N \times N$ identity matrix. The eigenvectors of an eigenvalue λ are the nonzero solutions of the linear homogeneous system $(\lambda I - P)X = 0$.

Google search : Markov chains and eigenvalues

Theorem (Perron–Frobenius theorem)

Let us consider an $N \times N$ transition matrix $P = (p_{ij})$ (i.e. $\forall i, j \in [\![1, N]\!]$, $p_{ij} \in [0, 1]$ and $\sum_{i=1}^{N} p_{ij} = 1$). Then :

Google search : Markov chains and eigenvalues

Theorem (Perron–Frobenius theorem)

Let us consider an $N \times N$ transition matrix $P = (p_{ij})$ (i.e. $\forall i, j \in [\![1, N]\!]$, $p_{ij} \in [0, 1]$ and $\sum_{i=1}^{N} p_{ij} = 1$). Then :

i) $\lambda = 1$ is one eigenvalue of P.

Google search : Markov chains and eigenvalues

Theorem (Perron–Frobenius theorem)

Let us consider an $N \times N$ transition matrix $P = (p_{ij})$ (i.e. $\forall i, j \in [\![1, N]\!]$, $p_{ij} \in [0, 1]$ and $\sum_{i=1}^{N} p_{ij} = 1$). Then :

i) $\lambda = 1$ is one eigenvalue of P.

ii) Any eigenvalue λ of P satisfies $|\lambda| \leq 1$.

Google search : Markov chains and eigenvalues

Theorem (Perron–Frobenius theorem)

Let us consider an $N \times N$ transition matrix $P = (p_{ij})$ (i.e. $\forall i, j \in [\![1, N]\!]$, $p_{ij} \in [0, 1]$ and $\sum_{i=1}^{N} p_{ij} = 1$). Then :

i) $\lambda = 1$ is one eigenvalue of P.

ii) Any eigenvalue λ of P satisfies $|\lambda| \leq 1$.

iii) There exists an eigenvector π of the eigenvalue 1, all the coordinates of which are greater or equal than zero. Without loss of generality we can suppose that the sum of its coordinates be 1.

Google search : Markov chains and eigenvalues

Theorem (Perron–Frobenius theorem)

Let us consider an $N \times N$ transition matrix $P = (p_{ij})$ (i.e. $\forall i, j \in [\![1, N]\!]$, $p_{ij} \in [0, 1]$ and $\sum_{i=1}^{N} p_{ij} = 1$). Then :

i) $\lambda = 1$ is one eigenvalue of P.

ii) Any eigenvalue λ of P satisfies $|\lambda| \leq 1$.

iii) There exists an eigenvector π of the eigenvalue 1, all the coordinates of which are greater or equal than zero. Without loss of generality we can suppose that the sum of its coordinates be 1.

$\Rightarrow$ **for the matrix associated with a web graph, there is always a stationary solution** π

For a proof, see any book of linear algebra

Google search : Markov chains and eigenvalues

Let us apply this theorem to the Google problem

Google search : Markov chains and eigenvalues

Let us apply this theorem to the Google problem

We make the following simplifying assumption :
P has a basis of eigenvectors $\mathcal{B} = \{v_1, \ldots, v_n\}$

Google search : Markov chains and eigenvalues

Let us apply this theorem to the Google problem

We make the following simplifying assumption :
P has a basis of eigenvectors $\mathcal{B} = \{v_1, \ldots, v_n\}$

We assume that v_1 is the π vector of Frobenius theorem.

Google search : Markov chains and eigenvalues

Let us apply this theorem to the Google problem

We make the following simplifying assumption :
P has a basis of eigenvectors $\mathcal{B} = \{v_1, \ldots, v_n\}$

We assume that v_1 is the π vector of Frobenius theorem.

$\forall v_i,\ \exists \lambda_i,\ P v_i = \lambda_i v_i$

Google search : Markov chains and eigenvalues

Let us apply this theorem to the Google problem

We make the following simplifying assumption :
P has a basis of eigenvectors $\mathcal{B} = \{v_1, \ldots, v_n\}$

We assume that v_1 is the π vector of Frobenius theorem.

$\forall v_i,\ \exists \lambda_i,\ Pv_i = \lambda_i v_i$

Let us take $X^t \neq (0, \ldots, 0)$ with $X^t = (x_1, \ldots, x_N)$ where $x_i \in [0, 1]$ and $\sum_{i=1}^N x_i = 1$ and decompose X in basis $\mathcal{B}$:
$X = \sum_{i=1}^N a_i v_i$

Google search : Markov chains and eigenvalues

Let us apply this theorem to the Google problem

We make the following simplifying assumption :
P has a basis of eigenvectors $\mathcal{B} = \{v_1, \ldots, v_n\}$

We assume that v_1 is the π vector of Frobenius theorem.

$\forall v_i,\ \exists \lambda_i,\ Pv_i = \lambda_i v_i$

Let us take $X^t \neq (0, ..., 0)$ with $X^t = (x_1, \ldots, x_N)$ where $x_i \in [0, 1]$ and $\sum_{i=1}^{N} x_i = 1$ and decompose X in basis $\mathcal{B}$:
$X = \sum_{i=1}^{N} a_i v_i$

We could show that $a_1 = 1$.

Google search : Markov chains and eigenvalues

Let us calculate PX :
$PX = \sum_{i=1}^{N} a_i P v_i = \sum_{i=1}^{n} a_i \lambda_i v_i,$
v_i being an eigenvector of the eigenvalueλ_i.

Google search : Markov chains and eigenvalues

Let us calculate PX :
$PX = \sum_{i=1}^{N} a_i P v_i = \sum_{i=1}^{n} a_i \lambda_i v_i,$
v_i being an eigenvector of the eigenvalueλ_i.

Iterating, we get : $P^n X = \sum_{i=1}^{n} a_i \lambda_i^n v_i$

Google search : Markov chains and eigenvalues

Let us calculate PX :
$PX = \sum_{i=1}^{N} a_i P v_i = \sum_{i=1}^{n} a_i \lambda_i v_i$,
v_i being an eigenvector of the eigenvalueλ_i.

Iterating, we get : $P^n X = \sum_{i=1}^{n} a_i \lambda_i^n v_i$

If $\forall \lambda_{i,\, i>1}, |\lambda| < 1$, then $\lim_{n\to\infty} P^n X = a_1 v_1 = \pi$

Google search : Markov chains and eigenvalues

Let us calculate PX :
$PX = \sum_{i=1}^{N} a_i P v_i = \sum_{i=1}^{n} a_i \lambda_i v_i$,
v_i being an eigenvector of the eigenvalueλ_i.

Iterating, we get : $P^n X = \sum_{i=1}^{n} a_i \lambda_i^n v_i$

If $\forall \lambda_{i,\, i>1}, |\lambda| < 1$, then $\lim_{n \to \infty} P^n X = a_1 v_1 = \pi$

$\rightarrow$ **this is what happens in our example !**

Google search : Markov chains and eigenvalues

Let us calculate PX :
$PX = \sum_{i=1}^{N} a_i P v_i = \sum_{i=1}^{n} a_i \lambda_i v_i$,
v_i being an eigenvector of the eigenvalueλ_i.

Iterating, we get : $P^n X = \sum_{i=1}^{n} a_i \lambda_i^n v_i$

If $\forall \lambda_{i,\, i>1}, |\lambda| < 1$, then $\lim_{n \to \infty} P^n X = a_1 v_1 = \pi$

$\rightarrow$ **this is what happens in our example !**

Remark : however the theorem does not guarantee that any matrix P satisfying the hypotheses of the theorem will have this property.

Google search : Markov chains and eigenvalues

Possible pathologies :

Google search : Markov chains and eigenvalues

Possible pathologies :

- the eigenvalue 1 may be a multiple root of the characteristic polynomial $\det(\lambda I - A) = 0$

Google search : Markov chains and eigenvalues

Possible pathologies :

- the eigenvalue 1 may be a multiple root of the characteristic polynomial $\det(\lambda I - A) = 0$
- The matrix P may have other eigenvalues λ than 1, with modulus equal to 1

Google search : Markov chains and eigenvalues

Possible pathologies :

- the eigenvalue 1 may be a multiple root of the characteristic polynomial $\det(\lambda I - A) = 0$
- The matrix P may have other eigenvalues λ than 1, with modulus equal to 1

We call a Markov transition matrix with no pathology a ***regular Markov transition matrix***

Google search : Markov chains and eigenvalues

Possible pathologies :

- the eigenvalue 1 may be a multiple root of the characteristic polynomial $\det(\lambda I - A) = 0$
- The matrix P may have other eigenvalues λ than 1, with modulus equal to 1

We call a Markov transition matrix with no pathology a ***regular Markov transition matrix***

Remedy : modify P to make it regular

Google search : Markov chains and eigenvalues

Possible pathologies :

- the eigenvalue 1 may be a multiple root of the characteristic polynomial $\det(\lambda I - A) = 0$
- The matrix P may have other eigenvalues λ than 1, with modulus equal to 1

We call a Markov transition matrix with no pathology a ***regular Markov transition matrix***

Remedy : modify P to make it regular

Let us consider the $N \times N$ matrix $Q = (q_{ij})$, with $\forall i, j \in [1, 2, \ldots, N], q_{ij} = \frac{1}{N}$

Google search : Markov chains and eigenvalues

Possible pathologies :

- the eigenvalue 1 may be a multiple root of the characteristic polynomial $\det(\lambda I - A) = 0$
- The matrix P may have other eigenvalues λ than 1, with modulus equal to 1

We call a Markov transition matrix with no pathology a ***regular Markov transition matrix***

Remedy : modify P to make it regular

Let us consider the $N \times N$ matrix $Q = (q_{ij})$, with $\forall i, j \in [1, 2, ..., N], q_{ij} = \frac{1}{N}$

We replace the matrix P of the web by $P_\beta = (1 - \beta)P + \beta Q$ for some small $\beta \in [0, 1]$

Google search : Markov chains and eigenvalues

Possible pathologies :

- the eigenvalue 1 may be a multiple root of the characteristic polynomial $\det(\lambda I - A) = 0$
- The matrix P may have other eigenvalues λ than 1, with modulus equal to 1

We call a Markov transition matrix with no pathology a ***regular Markov transition matrix***

Remedy : modify P to make it regular

Let us consider the $N \times N$ matrix $Q = (q_{ij})$, with $\forall i, j \in [1, 2, ..., N], q_{ij} = \frac{1}{N}$

We replace the matrix P of the web by $P_\beta = (1 - \beta)P + \beta Q$ for some small $\beta \in [0, 1]$

(Google uses $\beta = 0.15$)

Google search : Markov chains and eigenvalues

The following theorem guarantees that there exists a small β for which these pathologies vanish

Google search : Markov chains and eigenvalues

The following theorem guarantees that there exists a small β for which these pathologies vanish

Theorem

Given any Markov transition matrix P, there always exists a $\beta > 0$, as small as we wish, such that the matrix P_β is regular.

Google search : Markov chains and eigenvalues

The following theorem guarantees that there exists a small β for which these pathologies vanish

Theorem

Given any Markov transition matrix P, there always exists a $\beta > 0$, as small as we wish, such that the matrix P_β is regular.

Let π be the eigenvector of 1 for P_β, normalized so that the sum of its coordinates be 1.

Google search : Markov chains and eigenvalues

The following theorem guarantees that there exists a small β for which these pathologies vanish

Theorem
Given any Markov transition matrix P, there always exists a $\beta > 0$, as small as we wish, such that the matrix P_β is regular.

Let π be the eigenvector of 1 for P_β, normalized so that the sum of its coordinates be 1.

For P_β, given any nonzero vector X, where$X^t = (p_1, \dots, p_N)$ with $p_i \in [0, 1]$ and $\sum_{i=1}^{N} p_i = 1$, we have :

$$\lim_{n \to \infty} (P_\beta)^n X = \pi$$

Google search : Markov chains and eigenvalues

Calculation of the stationary distribution

$\rightarrow$ Not trivial when the matrix has billions of lines and columns

Google search : Markov chains and eigenvalues

Calculation of the stationary distribution

$\rightarrow$ Not trivial when the matrix has billions of lines and columns

$\rightarrow$ We use the previous theorem considering X_0 such that :

$$X_0^t = (\frac{1}{N}, \frac{1}{N}, \dots, \frac{1}{N})$$

Google search : Markov chains and eigenvalues

Calculation of the stationary distribution

$\rightarrow$ Not trivial when the matrix has billions of lines and columns

$\rightarrow$ We use the previous theorem considering X_0 such that :

$$X_0^t = (\frac{1}{N}, \frac{1}{N}, \ldots, \frac{1}{N})$$

We want to calculate $\lim_{n \to \infty} (P_\beta)^n X_0$

Google search : Markov chains and eigenvalues

Calculation of the stationary distribution

$\rightarrow$ Not trivial when the matrix has billions of lines and columns

$\rightarrow$ We use the previous theorem considering X_0 such that :

$$X_0^t = (\frac{1}{N}, \frac{1}{N}, \ldots, \frac{1}{N})$$

We want to calculate $\lim_{n \to \infty} (P_\beta)^n X_0$

In practice, $P_\beta^n X_0$ for some n between 50 and 100 usually provides us with a good approximation of π

Google search : Markov chains and eigenvalues

Calculation of the stationary distribution

$\rightarrow$ Not trivial when the matrix has billions of lines and columns

$\rightarrow$ We use the previous theorem considering X_0 such that :

$$X_0^t = (\frac{1}{N}, \frac{1}{N}, \ldots, \frac{1}{N})$$

We want to calculate $\lim_{n\rightarrow\infty}(P_\beta)^n X_0$

In practice, $P_\beta^n X_0$ for some n between 50 and 100 usually provides us with a good approximation of π

By induction, we calculate $X_{n+1} = P_\beta X_n$

Google search : Markov chains and eigenvalues

Calculation of the stationary distribution

$\rightarrow$ Not trivial when the matrix has billions of lines and columns

$\rightarrow$ We use the previous theorem considering X_0 such that :

$$X_0^t = (\frac{1}{N}, \frac{1}{N}, \dots, \frac{1}{N})$$

We want to calculate $\lim_{n\to\infty} (P_\beta)^n X_0$

In practice, $P_\beta^n X_0$ for some n between 50 and 100 usually provides us with a good approximation of π

By induction, we calculate $X_{n+1} = P_\beta X_n$

Remark : P_β has no zero entries, but most of the entries of P are zero.

Google search : Markov chains and eigenvalues

Let us take advantage of the previous remark :

$$X_{n+1} = (1 - \beta)PX_n + \beta QX_n$$

Google search : Markov chains and eigenvalues

Let us take advantage of the previous remark :

$$X_{n+1} = (1-\beta)PX_n + \beta QX_n$$

It is easy to verify that if X is a vector the sum of entries of which is 1, then $QX = X_0$

Google search : Markov chains and eigenvalues

Let us take advantage of the previous remark :

$$X_{n+1} = (1-\beta)PX_n + \beta QX_n$$

It is easy to verify that if X is a vector the sum of entries of which is 1, then $QX = X_0$

Hence, we can calculate the sequence

$$X_{n+1} = (1-\beta)PX_n + \beta X_0$$

□

Google search : Markov chains and eigenvalues

Conclusion

- elementary mathematics for a powerful tool

Google search : Markov chains and eigenvalues

Conclusion

- elementary mathematics for a powerful tool
- allows to find tricks to improve the ranking of your personal page by adding internal and external links in an optimal way

Google search : Markov chains and eigenvalues

Conclusion

- elementary mathematics for a powerful tool
- allows to find tricks to improve the ranking of your personal page by adding internal and external links in an optimal way
- one trick consists in replacing matrix Q by matrices reflecting the preferences of the web surfer

Google search : Markov chains and eigenvalues

Conclusion

- elementary mathematics for a powerful tool
- allows to find tricks to improve the ranking of your personal page by adding internal and external links in an optimal way
- one trick consists in replacing matrix Q by matrices reflecting the preferences of the web surfer

Remark : the Banach fixed point theorem can also be applied to this problem

Plan

Plan

Bibliography

Artificial Intelligence :

Artificial Intelligence, Stuart Russel and Peter Norvig. (pdf version available)

Bayesian Reasoning and Machine Learning David Barber (pdf version available)

http://web4.cs.ucl.ac.uk/staff/D.Barber/pmwiki/pmwiki.php?n=Brml.HomePage

Probabilities and Markov Chains :

Probabilités de l'Ingénieur. Variables aléatoires et simulation, Nicolas Bouleau, Hermann Éditeurs des Sciences et des Arts, 2002.

Bayesian networks :

Réseaux Bayésiens, Patrick Naïm et al., Éditions Eyrolles.

Bayesian networks, Judea Pearl, 1988, Handbook of Brain Theory and Neural Networks, MIT Press, pp. 157–160.