

21

Algorithmes des k-plus proches voisins

1. I. Introduction

 L'algorithme kNN (k-nearest neighbors) ou kppv (algorithme des k-plus proches voisins) est un algorithme (parmi les plus élémentaires) utilisé en intelligence artificielle.

 Note :

L'algorithme kNN est un **algorithme de classification et d'apprentissage supervisé** :

- Classification : on cherche à regrouper les individus de la population étudiée en classes «homogènes».
- Apprentissage supervisé : le programmeur ne donne pas de "critère explicite" pour la classification mais des données de référence déjà réparties dans les classes. C'est par «comparaison» avec ces données que l'ordinateur pourra classer les nouvelles données.

? L'idée générale est la suivante :

- On dispose d'une population statistique avec des mesures sur certains attributs des individus de cette population.
- Les individus se répartissent en plusieurs classes, plus ou moins bien caractérisées par les valeurs des attributs.
- Pour décider de la classe d'un nouvel individu, on compare ses attributs avec ceux des individus déjà classés.

Prenons l'exemple d'une population d'éléphants répartis en deux classes : les éléphants d'Afrique et les éléphants d'Asie.

Les éléphants d'Afrique ont des oreilles plus grandes que les éléphants d'Asie, l'éléphant d'Afrique est plus grand que l'éléphant d'Asie.

Les attributs utilisés pourraient donc être :

- taille de l'éléphant,
- taille des oreilles.

Arrive un nouvel individu.

On ne sait pas a priori de quelle classe il relève. Pour le décider, on décide de comparer les valeurs de ses attributs (hauteur de l'éléphant, taille des oreilles) à ceux de la population déjà répartie en classe.

Le problème est que, parmi les éléphants d'Asie il y en a des plus gros (qui se rapprochent donc de l'éléphant d'Afrique) ou avec de plus grandes oreilles. De même les éléphants d'Afrique ne sont pas uniformes, il en existe

des plus petits, etc...

Bref, les attributs choisis ne permettent pas nécessairement de conclure de façon évidente. On va donc comparer notre nouvel individu à tous les individus présents. Et par exemple retenir les k=3 éléphants ayant des attributs les "plus proches" de notre individu (on a choisi k=3 mais on peut choisir k=4 ou une autre valeur de k).

Si parmi les k individus les plus proches de notre nouvel arrivant, il y a plus d'éléphants d'Afrique que d'éléphants d'Asie, on décide qu'il s'agit d'un éléphant d'Afrique.

La décision n'est évidemment pas absolument fiable (et il peut arriver qu'en changeant la valeur de k, on change de conclusion!) mais avec une population de départ suffisamment importante et quelques heuristiques pour le choix de k, le principe donne de bons résultats dans un certain nombre de contextes.

2. II. Activité d'introduction

On peut	In [3]	On peut	In [4]:	QuestionRéponsePrenons	In [5]	Question	In [6]	In [7]
mémoriser		alors	Out[4]:	1 : Quel : le premier		2 : Ecrivez		
les		afficher	[{ o	est le		les	Out[6]:	' a
données		le	m' type de	Pokemon		instructions		u
du fichier		contenu	' données	de la		permettant		
csv dans		de la	li ga des	table		d'accéder		
une table		table	tu éléments	(n'oubliez		à son type,		
pokemons		pokemo	r' du	pas		puis à son		
afin de		ns .	' tableau	d'exécuter		nombre de		
pouvoir			yp pokemon	la cellule		points de		
travailler			e' s ?	ci-		vie et enfin		
sur les			u'	dessous		à sa valeur		
données.			oi	pour que		d'attaque.		
			nt	la variable				
			s	pokemon1				
			de	créée).				
			vi					
			e'					
			'					
			5'					
			'					
			tt					
			aq					
			u					
			e'					
			'					
			0					
			5'					
				{				
				o				
				m'				
				'				
				ar				
				ga				
				nt				
				u				
				a'				
				'				
				yp				
				e'				
				'				
				a				
				u'				
				'				

oi
nt
s
de
vi
e'
'
θ'
'
tt
aq
u
e'
'
2'

{
o
m'
'
ar
ab
af
f
e'
'
yp
e'
'
a
u'
'
oi
nt
s
de
vi
e'
'
9'
'
tt
aq
u
e'
'
3'

{
o
m'
'
la
mi
ra
l'
'
yp
e'
'
a
u'
'
oi
nt
s
de
vi
e'
'
5'
'
tt
aq
u
e'
'

θ
θ'

{
o
m'
,

re
fa
de
t'
,

yp
e'
,

s
y'
,

oi
nt
s
de
vi
e'
,

5'
,

tt
aq
u
e'
,

2
5'

{
o
m'
,

ro
cr
od
i
l'
,

yp
e'
,

a
u'
,

oi
nt
s
de
vi
e'
,

5'
,

tt
aq
u
e'
,

θ'

{
o
m'
,

eo
xy
s'
,

yp
e'

'
s
y'
'
oi
nt
s
de
vi
e'
'
θ'
'
tt
aq
u
e'
'
θ'

{
o
m'
'
eo
xy
s'
'
yp
e'
'
s
y'
'
oi
nt
s
de
vi
e'
'
θ'
'
tt
aq
u
e'
'
5'

{
o
m'
'
eo
xy
s'
'
yp
e'
'
s
y'
'
oi
nt
s
de
vi
e'
'
θ'
'
tt
aq
u
e'

'
5
θ'

{
o
m'
,

eo
xy
s'
,

yp
e'
,

s
y'
,

oi
nt
s
de
vi
e'
,

θ'
,

tt
aq
u
e'
,

8
θ'

{
o
m'
,

ca
yo
n'
,

yp
e'
,

a
u'
,

oi
nt
s
de
vi
e'
,

9'
,

tt
aq
u
e'
,

9'

{
o
m'
,

ok
o'
,

yp
e'
,

s
y'

'
oi
nt
s
de
vi
e'
'
5'
'
tt
aq
u
e'
'
θ'

{
o
m'
'
am
bl
as
t'
'
yp
e'
'
a
u'
'
oi
nt
s
de
vi
e'
'
1'
'
tt
aq
u
e'
'
3'

{
o
m'
'
ob
o
u'
'
yp
e'
'
a
u'
'
oi
nt
s
de
vi
e'
'
θ'
'
tt
aq
u
e'
'
θ'

{
o
m'
'
ro
re
t'
'
yp
e'
'
s
y'
'
oi
nt
s
de
vi
e'
'
θ'
'
tt
aq
u
e'
'
5'

{
o
m'
'
at
el
ou
tr
e'
'
yp
e'
'
a
u'
'
oi
nt
s
de
vi
e'
'
5'
'
tt
aq
u
e'
'
5'

{
o
m'
'
es
me
re
ll
a'
'
yp
e'
'
s

y'
,

oi

nt

s

de

vi

e'

,

θ'

,

tt

aq

u

e'

,

5'

{

o

m'

,

e

w'

,

yp

e'

,

s

y'

,

oi

nt

s

de

vi

e'

,

θ

θ'

,

tt

aq

u

e'

,

θ

θ'

{

o

m'

,

ew

tw

o'

,

yp

e'

,

s

y'

,

oi

nt

s

de

vi

e'

,

θ

6'

,

tt

aq

u

e'

'
1
θ'
{
o
m'
'
ew
tw
o'
'
yp
e'
'
s
y'
'
oi
nt
s
de
vi
e'
'
θ
6'
'
tt
aq
u
e'
'
5
θ'
{
o
m'
'
ew
tw
o'
'
yp
e'
'
s
y'
'
oi
nt
s
de
vi
e'
'
θ
6'
'
tt
aq
u
e'
'
9
θ'
{
o
m'
'
un
n
a'
'
yp

e'
,

s
y'

,

oi
nt

s
de

vi
e'

,

6'
,

tt
aq

u
e'

,

5'
,

{
o

m'

,

uc
le

o
s'

,

yp
e'

,

s
y'

,

oi
nt

s
de

vi
e'

,

5'
,

tt
aq

u
e'

,

θ'
,

{
o

m'

,

ct
il

le
r

y'

,

yp
e'

,

a
u'

,

oi
nt

s
de

vi
e'

,

5'
,

tt
aq
u
e'
'
θ
5'

{
o
m'
'
ke
ok
e'
'
yp
e'
'
s
y'
'
oi
nt
s
de
vi
e'
'
5'
'
tt
aq
u
e'
'
3'

{
o
m'
'
hi
on
e'
'
yp
e'
'
a
u'
'
oi
nt
s
de
vi
e'
'
θ'
'
tt
aq
u
e'
'
θ'

{
o
m'
'
oi
ss
or
o
y'

'
yp
e'
'
a
u'
'
oi
nt
s
de
vi
e'
'
θ'
'
tt
aq
u
e'
'
2'

{
o
m'
'
ri
np
lo
u
f'
'
yp
e'
'
a
u'
'
oi
nt
s
de
vi
e'
'
4'
'
tt
aq
u
e'
'
6'

{
o
m'
'
os
ab
ys
s'
'
yp
e'
'
a
u'
'
oi
nt
s
de
vi
e'
'

5'
'
tt
aq
u
e'
'
4'
{
o
m'
'
id
er
el
l
a'
'
yp
e'
'
s
y'
'
oi
nt
s
de
vi
e'
'
θ'
'
tt
aq
u
e'
'
5'
{
o
m'
'
po
in
k'
'
yp
e'
'
s
y'
'
oi
nt
s
de
vi
e'
'
θ'
'
tt
aq
u
e'
'
5'
{
o
m'
'
ym
bi

o
s'
'
yp
e'
'
s
y'
'
oi
nt
s
de
vi
e'
'
1
θ'
'
tt
aq
u
e'
'
5'

{
o
m'
'
ar
pa
u
d'
'
yp
e'
'
a
u'
'
oi
nt
s
de
vi
e'
'
θ'
'
tt
aq
u
e'
'
5'

{
o
m'
'
ip
lo
u
f'
'
yp
e'
'
a
u'
'
oi
nt
s
de
vi

e'
,

3'
,

tt

aq

u

e'
,

1'