DJILLALI LIABES UNIVERSITY OF SIDI BEL ABBES
FACULTY OF EXACT SCIENCES
DEPARTMENT OF COMPUTER SCIENCES

Module : Apprentissage Automatique
1ST YEAR OF MASTER'S DEGEREE IN
NETWORKS,SYSTEMS & INFORMATION SECURITY(RSSI)
2021/2022

Solution of TD4 K- Nearest Neighbor

Author:
HADJAZI Mohammed
Hisham
Group: 01 / RSSI

Supervisor:
Pr.ELBERRICHI Zakaria

November 12, 2021

Contents

Chapter 1

Fiche TD-04 Solutions

1.1 Questions de cours.

1.1.1 1. Critiquer l'algorithme KNN (ses limites).

I think the biggest limit to KNN algorithm is the calculation time as it requires a lot of work and even if done by machines it will be limited to data science applications and not useful in any real time applications where we require the results immediately. even for data science applications if the dataset is large it will be very costly as it requires the dataset to be loaded to the memory.

Another problem with This algorithm is its sensitivity to scales as sometimes one feature can overpower the rest if it has a larger value, although this problem can be solved with normalization it still can sometimes play as a factor plus it will consume more time to normalize.

1.1.2 2. Si on voulait utiliser l'algorithme Knn pour une régression, comment procéderait-on ?

One of KNN biggest advantages is that it can be used for both classification and regression problems as it uses **feature similarity** to predict the values of any new data points.

In regression it works the same way as we calculate exactly the same way after that the results we get from the distances and group the results that are close to each other on groups.

1.1.3 3. Principe de K-NN : dis-moi, qui sont tes voisins, je te dirais qui tu es ! Est-il toujours vrai ?

yes it works if you choose the K value wisely, and with several trails as chooosing a high number can result in **Underfitting** while a small number will result in **Overfitting** but with the right number it can give very good results of-course considering that our dataset is of high quality.

1.1.4 4. Pour quel type d'applications Knn donne de bon resultats (rechercher sur le net) ?

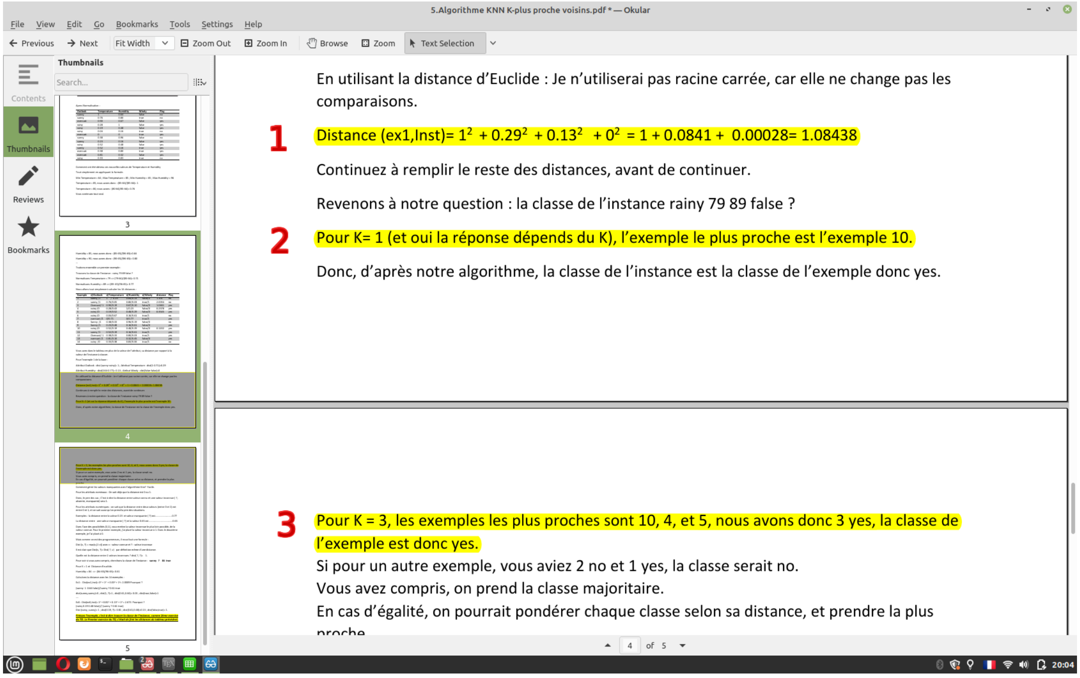
1. **Agriculture**
2. **Facial recognition**
3. **Finance**

4. **Medical**
5. **Recommendation systems** Amazon, Hulu, Netflix, etc
6. **Text mining**

[1]

1.2 Exercice 1 : Finir la colonne distance du 1 er tableau. Vérifier que nos conclusions sur la classe étaient correctes.

Example	d/Outlook	d/Temperature	d/Humidity	d/Windy	distance	Play
1	sunny /1	1 / 0.29	0.64/0.13	false/0	1.101	no
2	sunny /1	0.76/0.05	0.80/0.03	true/1	2.0034	no
3	Overcast/ 1	0.90/0.19	0.67/0.10	false/0	1.0461	yes
4	rainy /0	0.28/0.43	1/0.23	false/0	0.2378	yes
5	rainy /0	0.19/0.52	0.48/0.29	false/0	0.3545	yes
6	rainy /0	0.04/0.67	0.16/0.61	true/1	1.821	no
7	overcast /1	0/0.71	0/0.77	true/1	3.097	yes
8	Sunny /1	0.38/0.33	0.96/0.19	false/0	1.145	no
9	Sunny /1	0.23/0.48	0.16/0.61	false/0	1.6025	yes
10	rainy /0	0.52/0.19	0.48/0.29	false/0	0.1202	yes
11	sunny /1	0.52/0.19	0.16/0.61	true/1	2.4082	yes
12	Overcast/ 1	0.38/0.33	0.80/0.03	true/1	2.1098	yes
13	overcast /1	0.81/0.10	0.32/0.45	false/0	1.2125	yes
14	rainy /0	0.33/0.38	0.83/0.06	true/1	1.148	no


5.Algorithme KNN K-plus proche voisins.pdf * — Okular
File View Edit Go Bookmarks Tools Settings Help
Previous Next Fit Width Zoom Out Zoom In Browse Zoom Text Selection
Thumbnails
Search...
Contents
Thumbnails
Reviews
Bookmarks
3
4
5
4 of 5
20:04

En utilisant la distance d'Euclide : Je n'utiliserai pas racine carrée, car elle ne change pas les comparaisons.

1 Distance (ex1,Inst)= 1^2 + 0.29^2 + 0.13^2 + 0^2 = 1 + 0.0841 + 0.00028= 1.08438

Continuez à remplir le reste des distances, avant de continuer.

Revenons à notre question : la classe de l'instance rainy 79 89 false ?

2 Pour K= 1 (et oui la réponse dépends du K), l'exemple le plus proche est l'exemple 10.

Donc, d'après notre algorithme, la classe de l'instance est la classe de l'exemple donc yes.

3 Pour K = 3, les exemples les plus proches sont 10, 4, et 5, nous avons donc 3 yes, la classe de l'exemple est donc yes.
Si pour un autre exemple, vous aviez 2 no et 1 yes, la classe serait no.
Vous avez compris, on prend la classe majoritaire.
En cas d'égalité, on pourrait pondérer chaque classe selon sa distance, et prendre la plus proche

FIGURE 1.1: Errors Screen Shot

There are 3 errors in the cours the first one is no:1 in picture $Distance(ex1, Inst) = 1^2 + 0.29^2 + 0.13^2 + 0^2 = 1.101$.

The second error is shown in no:2 **for k=1 , the closest instance is 3 with a distance of 1.0461 with a class 'yes'**.

The third error comes in no:3 **As for k=3 The nearest neighbours are 1,3,8 with classes 'no', 'yes', 'no' to favor 'no'**.

1.3 Exercice 2 : Faire un tableau et trouver la classe du dernier exemple (avec valeur manquante).

Im not sure I understood the question as what table to create ? but to find the class of the last instance in the cours which is **sunny, ? , 84 , true** ?

$Distance(ex1, Inst) = 1^2 + 1^2 + 0.61^2 + 1^2 = 3.3721$

so for k=1 the closest is 7 which is class 'yes'.

and for k=3 we get 7,11,12 for 'yes', 'yes', 'yes'.

1.4 Exercice 3 : En utilisant l'algorithme de Knn (k = 1, puis K = 3), la distance de Manhattan, puis la distance d'Euclide, classer les instances : 1,2,3

1. **N : ? : M : 85**
2. **O : 49 : ? : 110**
3. **O : 55 ; F ; 100**

Etude	Age	Sexe	Prêt	Rembourse ?
O	25	M	40	O
N	35	M	60	O
N	45	F	?	O
?	20	F	20	N
N	35	M	120	N
O	52	F	18	N
O	23	M	95	O
O	?	M	62	O
N	60	F	100	N
O	48	F	220	O
O	33	?	150	O

1.4.1 Instance 1

MODIFIED TABLE FOR INSTANCE 1 with calculated Distances

Etude		**Age**		**Sexe**		**Prêt**		**Distance**	**Rembourse ?**
O	0	25	0.125	M	1	40	0.108910891	1.027486582	O
N	1	35	0.375	M	1	60	0.207920792	2.183856056	O
N	1	45	0.625	F	0	?	0.604	1.755441	O
?	1	20	0	F	0	20	0.00990099	1.00009803	N
N	1	35	0.375	M	1	120	0.504950495	2.395600002	N
O	0	52	0.8	F	0	18	0	0.64	N
O	0	23	0.075	M	1	95	0.381188119	1.150929382	O
O	0	?	0.875	M	1	62	0.217821782	1.813071329	O
N	1	60	1	F	0	100	0.405940594	2.164787766	N
O	0	48	0.7	F	0	220	1	1.49	O
O	0	33	0.325	?	1	150	0.653465347	1.532641959	O

Instance 1 is (O, 55, F, 100) therefore $Distance(Inst1) = 0^2 + 0.875^2 + 0^2 + 0.4^2 = 0.925$

Result for K=1 is 4 = 1.00009803 which is class of Rembourse = 'N'

Result for K=3 is 1,4,7 which is class of 'O', 'N', 'O' to Rembourse = 'O'

1.4.2 Instance 2

MODIFIED TABLE FOR INSTANCE 2 with calculated Distances

Etude		**Age**		**Sexe**		**Prêt**		**Distance**	**Rembourse ?**
O	1	25	0	M	0	40	0.108910891	1.011861582	O
N	0	35	0	M	0	60	0.207920792	0.043231056	O
N	0	45	0	F	1	?	0.604	1.364816	O
?	1	20	0	F	1	20	0.00990099	2.00009803	N
N	0	35	0	M	0	120	0.504950495	0.254975002	N
O	1	52	0	F	1	18	0	2	N
O	1	23	0	M	0	95	0.381188119	1.145304382	O
O	1	?	0	M	0	62	0.217821782	1.047446329	O
N	0	60	0	F	1	100	0.405940594	1.164787766	N
O	1	48	0	F	1	220	1	3	O
O	1	33	0	?	1	150	0.653465347	2.427016959	O

Instance 2 is (O, 55, F, 100) therefore $Distance(Inst2) = 0^2 + 0.^2 + 0^2 + 0.33^2 =$ 0.1089

Result for K=1 is 2 = 0.043231056 which is class of Rembourse = 'O'

Result for K=3 is 1,2,5 which is class of 'O', 'O', 'N' to Rembourse = 'O'

1.4.3 Instance 3

MODIFIED TABLE FOR INSTANCE 3 with calculated Distances

Etude		Age		Sexe		Prêt		Distance	Rembourse ?
O	0	25	0.125	M	0	40	0.108910891	0.027486582	O
N	1	35	0.375	M	0	60	0.207920792	1.183856056	O
N	1	45	0.625	F	0	?	0.604	1.755441	O
?	1	20	0	F	0	20	0.00990099	1.00009803	N
N	1	35	0.375	M	0	120	0.504950495	1.395600002	N
O	0	52	0.8	F	0	18	0	0.64	N
O	0	23	0.075	M	0	95	0.381188119	0.150929382	O
O	0	?	0.875	M	0	62	0.217821782	0.813071329	O
N	1	60	1	F	0	100	0.405940594	2.164787766	N
O	0	48	0.7	F	0	220	1	1.49	O
O	0	33	0.325	?	0	150	0.653465347	0.532641959	O

Instance 3 is (O, 49, ?, 110) therefore $Distance(Inst3) = 0^2 + 0.725^2 + 0^2 + 0.45^2 =$ 0.728125

Result for K=1 is 8 = 0.813071329 which is class of Rembourse = 'O'

Result for K=3 is 4,6,8 which is class of 'N', 'N', 'O' for Rembourse = 'N'

Bibliography

[1] Arman Hussain. *K-Nearest Neighbors (KNN) and its applications*. July 2020. URL: https://medium.com/@arman_hussain786/k-nearest-neighbors-knn-and-its-applications-7891a4a916c6.