

M2 ISL Sécurité Appliquée- TD1.

k -anonymat et apprentissage supervisé

Jean-François COUCHOT
couchot@arobase.femto-st.fr

13 novembre 2022

Exercice 0.1 (Rappels sur le k -anonymat sur un micro exemple). On considère le jeu de données D de la table 1, inspiré de ¹.

- Proposer une version 2-anonyme D_{k2}^g en considérant les généralisations suivantes :
 - Statut marital : séparé(e), célibataire, veuf/veuve, célibataire \rightsquigarrow *
 - Age : 20,23,24 \rightsquigarrow [20,24], 25,28,29 \rightsquigarrow [25,29], \rightsquigarrow *
 - CP : 32021,32024,32027 \rightsquigarrow 3202*, 32042,32045,32046 \rightsquigarrow 3204*, \rightsquigarrow *
- Calculer le Loss sur D_{k2}^g .

	ID	QID			sensible
#	Nom	stat. conjugal	Age	CP	Crime
1	Joe	Séparé	29	32042	Meurtre
2	Jill	Célibataire	20	32021	Vol
3	Sue	Veuve	24	32024	Trafic
4	Abe	Séparé	28	32046	Agression
5	Bob	Veuf	25	32045	Piratage
6	Amy	Célibataire	23	32027	Indécence

TABLE 1 – Données de criminalité

Exercice 0.2 (Apprentissage bayésien naïf). On considère le tableau suivant (extrait de ²) où, pour chacune des 14 personnes on connaît les valeurs de trois attributs Att1, Att2 et Att3. A partir de ces attributs, on cherche à savoir à quelle classe appartient la personne.

N	Att1	Att2	Att3	Classe	N	Att1	Att2	Att3	Classe
1	A	L	T	C1	8	B	L	T	C1
2	A	H	T	C2	9	B	H	F	C1
3	A	H	F	C2	10	C	H	T	C2
4	A	H	F	C2	11	C	L	T	C2
5	A	L	F	C1	12	C	H	F	C1
6	B	H	T	C1	13	C	H	F	C1
7	B	H	F	C1	14	C	H	F	C1

- Construire le classifieur bayésien naïf correspondant
- Prédire les classes des exemples suivants :

N	Att1	Att2	Att3
15	A	H	F
16	B	L	T
17	C	L	F

1. , V., McDonagh, P., Cerqueus, T., & Murphy, L. (2014). A systematic comparison and evaluation of k-anonymization algorithms for practitioners. Transactions on data privacy, 7(3), 337-370.

2. http://abdelhamid-djeffal.net/web_documents/tdbayes.pdf