

# M2 ISL Sécurité Appliquée- TD1.

## Apprentissages supervisés et $k$ -anonymat.

Jean-François COUCHOT

[couchot@arobase.femto-st.fr](mailto:couchot@arobase.femto-st.fr)

6 novembre 2023

### 1 Apprentissages supervisés

**Exercice 1.1 (Apprentissage bayésien naïf).** On considère le tableau suivant (extrait de<sup>1</sup>) où, pour chacune des 14 personnes on connaît les valeurs de trois attributs  $X_1$ ,  $X_2$  et  $X_3$ . A partir de ces attributs, on cherche à savoir à quelle  $Y$  appartient la personne.

$N$	$X_1$	$X_2$	$X_3$	$Y$	$N$	$X_1$	$X_2$	$X_3$	$Y$
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

1. Construire le classifieur bayésien naïf correspondant
2. Prédire les classes des exemples suivants :

$N$	$X_1$	$X_2$	$X_3$
15	C	L	F
16	B	L	T
17	A	H	F

**Exercice 1.2 (Régression Multiple par Moindres carrés ordinaires).**

Cet exercice est inspiré de<sup>2</sup>. On considère 5 observations en 2 dimensions, étiquetées par  $y$  :

$x_1$	0.12	0.14	0.31	0.37	0.49
$x_2$	29.0	33.0	17.0	21.0	12.0
$y$	21	24.3	12.7	15.6	9

1. Exprimer le problème à l'aide d'un système matriciel.
2. Montrer qu'une solution approximative de la fonction objectif est  $\hat{\omega} = (-0.577, 1.499, 0.744)$ . On pourra donner le code python qui effectue le calcul.
3. Que donnerait votre prédicteur pour des valeurs  $(x_1, x_2) = (0.25, 30.0)$  ?
4. Evaluer la somme des résidus quadratiques.

1. [http://abdelhamid-djeffal.net/web\\_documents/tdbayes.pdf](http://abdelhamid-djeffal.net/web_documents/tdbayes.pdf)

2. <https://www.dunod.com/sciences-techniques/introduction-au-machine-learning-1>

## 2 $k$ -anonymat

**Exercice 2.1 (2-anonymat sur un micro exemple).** On considère le jeu de données  $D$  de la table 1, inspiré de<sup>3</sup>.

1. Proposer une version 2-anonyme  $D_{k2}^g$  en considérant les généralisations suivantes :
  - Statut marital : {séparé(e), célibataire, veuf/veuve}  $\rightsquigarrow$  \*
  - 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\*,  $\rightsquigarrow$  \*
2. 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

3. , 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.