Sécurité Appliquée-TP 4
Protection de la vie privée-PVP

Jean-François COUCHOT
couchot[arobase]femto-st[point]fr

24 novembre 2021

On considère à nouveau le jeu de données sur le diabète comme au TP1.

Exercice 0.1 (Apprentissage bayesien naïf avec `sklearn`).

1. *Récupérer le jeu de données.*
2. *S'inspirer du tutoriel pour classifier les personnes à l'aide d'un apprentissage bayésien naïf en exploitant la bibliothèque* `sklearn`.
3. *Vérifier que la précision de la prédiction ne change pas à chaque exécution.*
4. *Quelle serait la prédiction pour des personnes avec les caractéristiques ci-dessous ? Vérifier avec les données originales.*
   [7, 184, 84, 33, 0, 35.5, 0.355, 41]
   [6, 109, 60, 27, 0, 25.0, 0.206, 27]

Exercice 0.2 (Apprentissage bayesien naïf respectueux avec `diffprivlib`).
Maintenue par IBM, `DiffprivLib` *est une bibliothèque en protection de la vie privée à base de confidentialité différentielle.*

1. *Installer cette bibliothèque.*
2. *En s'inspirant du site github de diffprivlib reprendre l'exercice précédent.*
3. *Vérifier que la précision de la prédiction varie à chaque exécution même pour la valeur d'epsilon par défaut égale à 1.*
4. *Faire varier epsilon dans* $[0.001, 0.01, 0.1, 1, 10, 100]$ *et constater que les prédiction sont de plus en plus précises. Vers quelle valeur de précision cela converge-t-il ?*
5. *Quelle serait ici la prédiction pour des personnes avec les caractéristiques ci-dessous ?*
   [7, 184, 84, 33, 0, 35.5, 0.355, 41]
   [6, 109, 60, 27, 0, 25.0, 0.206, 27]
6. *Construire un graphique affichant la valeur moyenne de précision pour 10 prédictions pour* ϵ *variant dans* $[0.001, 0.01, 0.1, 1, 10, 50, 100]$.