M1-ISL
Option Initiation à la Recherche.
Projet d'évaluation théorique et expérimentale d'une démarche de PVP

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

25 janvier 2022

1 Introduction

La méthode de réponse randomisée de Warner [War65] vue en cours est un mécanisme qui considère une valeur binaire (dans $\{\text{OUI}, \text{NON}\}$) et qui retourne une réponse dans cette ensemble en introduisant du hasard.

L'algorithme de réponse randomisée généralisée de Kairouz [KBR16], nommé $\mathcal{KRR}$ par la suite, est une adaptation de ce mécanisme à la situation où la valeur n'est plus binaire mais appartient à un ensemble $\{v_1, \ldots, v_k\}$, où toutes les valeurs v_1,...,v_k sont différentes (c'est un ensemble à k valeurs).

Étant donné un utilisateur dont la réponse serait $v_i \in \{v_1, \ldots, v_k\}$. Le mécanisme $\mathcal{KRR}$ va retourner une valeur dans $\{v_1, \ldots, v_k\}$, selon les probabilité suivantes :

$$\Pr[\mathcal{KRR}(v_i) = v_i] = \frac{e^\epsilon}{k - 1 + e^\epsilon} \quad (1)$$

$$\Pr[\mathcal{KRR}(v_i) = v_j] = \frac{1}{k - 1 + e^\epsilon} \text{ avec } i \neq j \quad (2)$$

Le première probabilité donnée à l'équation (1) est celle de conserver la donnée (v_i reste v_i). L'équation (2) dit juste avec quelle probabilité on passe d'une valeur v_i à une valeur v_j différente.

Ici ϵ est un paramètre de l'algorithme que l'on peut choisir parmi les réels positifs, par exemple dans l'ensemble $\{10^{-2}, 5 * 10^{-2}, 10^{-1}, 0.5, 1, 5, 10\}$. C'est un mécanisme qui vérifie la confidentialité différentielle (on verra ceci en M2). Comme pour tous les mécanismes vérifiant cette propriété, plus ϵ est petit, plus l'ajout de bruit est important, meilleure est la protection de la vie privée, et inversement.

Comme le mécanisme de réponse randomisée, $\mathcal{KRR}$ permet de protéger individuellement chaque utilisateur en modifiant de manière aléatoire sa réponse. De même, étant donné un grand nombre de réponses, il permet d'obtenir des statistiques globales grâce à un estimateur.

L'objectif de ce projet est d'évaluer la qualité de ce mécanisme au travers de son estimateur et des prédictions réalisées avec et sans ce mécanisme.

Le travail est à réaliser par groupe de 2 ou 3 au plus. Il s'agira de :

— rendre un rapport pour les justifications théoriques ;
— réaliser un notebook sur Google Colab que vous partagerez au moyen d'un lien avec moi ; plus il sera complet (code + commentaires + courbes), mieux cela sera ;
— le tout pour le 11/02/2022 à 23h45.

Voici un plan que l'on peut suivre.

2 Statistiques sur l'estimateur

Exercice 2.1. *On considère un ensemble de N personnes, chacune ayant une réponse personnelle dans $\{v_1, \ldots, v_k\}$. Soit f_i le nombre de fois où la valeur v_i apparaît dans l'ensemble initial de réponses personnelles.*

Soit r_i le nombre de fois où la valeur v_i apparaît dans l'ensemble obtenu après application du mécanisme $\mathcal{KRR}$ à chaque réponse individuelle.

1. *Montrer que $\hat{f}_i$ défini par*

$$\hat{f}_i = \frac{k-1+e^\epsilon}{e^\epsilon - 1} r_i - \frac{N}{e^\epsilon - 1} \tag{3}$$

est un estimateur non biaisé de f_i.

2. *Montrer que la variance de $\hat{f}_i$, noté $Var[\hat{f}_i]$ est définie par*

$$Var[\hat{f}_i] = \left(\frac{k-1+e^\epsilon}{e^\epsilon - 1}\right)^2 Var[\hat{r_i}] \tag{4}$$

$$Var[\hat{r_i}] = f_i \times \delta_1(1-\delta_1) + (N - f_i) \times \delta_0(1-\delta_0) \tag{5}$$

Exercice 2.2 (Estimateur : implantation et vérification). *On considère le jeu de données adult de l'UCI déjà vu en TD et particulièrement l'attribut* `marital-status`.

1. *Afficher les fréquences des différentes valeurs de cet attribut.*
2. *Implanter le mécanisme $\mathcal{KRR}$ et l'estimateur $\hat{f}$.*
3. *Appliquer le mécanisme $\mathcal{KRR}$ à chaque réponse avec une valeur de ϵ fixée à 1.*
4. *Vérifier que l'estimateur $\hat{f}$ permet de retrouver des fréquences similaires à celles de la question 1.*

Exercice 2.3 (Estimateur : statistiques). *Toujours sur le même jeu de données adult de l'UCI et le même attribut* `marital-status`*, contruire des statistiques de dispersion de cet estimateur.*

3 Application du mécanisme $\mathcal{KRR}$ à du Machine Learning

L'objectif ici est d'évaluer l'application du mécanisme $\mathcal{KRR}$ comme un prétraitement pour du Machine learning, celui de forêt alatoire, par exemple.

On rappelle que $\mathcal{KRR}$ ajoute du bruit à des valeurs dans un ensemble. Dans ce qui suit, on considérera donc les attributs discrets ['workclass', 'education', 'occupation', 'relationship', 'race', 'sex','native-country'] et l'on essaiera de deviner la valeur de 'marital-status'.

Exercice 3.1 (Sélection des attributs et valeur de baseline).

1. *A partir du jeu de données complet, construire un dataset où* `X` *et* `y` *ne contiennent que les attributs discrets à utiliser pour l'apprentissage et la valeur à prédire.*
2. *Découper "honnêtement"* `X` *et* `y` *en* `X_train`, `X_test`, `y_train` *et* `y_test`.
3. *Evaluer l'algorithme de forêt alatoire sur ce jeu de données. En déduire une valeur de baseline.*

Exercice 3.2 (Evaluation d'un apprentissage de données bruitées par le mécanisme $\mathcal{KRR}$).

1. *Montrer théoriquement que si l'on veut évaluer le mécanisme $\mathcal{KRR}$, il faut*
   (a) *appliquer le mécanisme $\mathcal{KRR}$ sur* `X_train`, `X_test` *et* `y_train` *et obtenir respectivement ;* `Xp_train`, `Xp_test` *et* `yp_train`
   (b) *faire l'apprentissage sur le couple (*`Xp_train`, `yp_train`*);*
   (c) *prédire les réponses correspondantes à* `Xp_test` *;*
   (d) *comparer celles-ci à* `y_test` *qui n'a pas été modifié.*
2. *Mettre en place cette démarche et afficher une valeur de précision.*

Exercice 3.3 (Analyse statistiques d'un apprentissage de données bruitées par le mécanisme $\mathcal{KRR}$).

1. *Répéter un grand nombre de fois la mesure de précision dont la démarche est détaillée à l'exercice précédent.*
2. *Construire des statistiques de moyenne de précision et de dispersion.*
3. *Etendre cette étude en faisant varier ϵ dans l'ensemble $\{10^{-2}, 5*10^{-2}, 10^{-1}, 0.5, 1, 5, 10\}$.*
4. *Illustrer ceci au moyen de courbes.*

Références

[KBR16] Peter Kairouz, Keith Bonawitz, and Daniel Ramage. Discrete distribution estimation under local privacy. In *International Conference on Machine Learning*, pages 2436–2444. PMLR, 2016.

[War65] Stanley L Warner. Randomized response : A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309) :63–69, 1965.