

HW9_Programming: The Experiment of t-closeness

Group: yayaya

108753109 資科碩一 陳庭軒

108753134 資科碩一 高士傑

108753124 資科碩一 楊芝辰

108753102 資科碩一 吳映函

1. Dataset: Adult (test size =0.3×dataset)

2. Programming Language: Python

3. Simple report:

(1) How do you perturb the data?

A: By t-closeness.

如果等價類 E 中的敏感屬性取值分佈與整張表中該敏感屬性的分佈的距離不超過閾值 t ，則稱 E 滿足 t-closeness。

如果數據表中所有等價類都滿足 t-closeness，則稱該表滿足 t-closeness。

(2) What is your machine learning model used in this experiment?

A: 機器學習模型使用 Python sklearn 底下的 svm.SVC。SVC 有很多參數，例如：gamma 值、probability 等，在這次實驗中所有參數皆使用預設值。

(3) Effectiveness measure: Misclassification Error, Accuracy, Precision, Recall, and AUC

A: 預測結果為 Salary (Salary > 50K 或 Salary <= 50K)

Misclassification Error: 0.18332145402708477

Accuracy: 0.7892150841419973

Confusion Matrix:

混淆矩陣

```
[[5729 430]
 [1286 696]]
```

Classification Report (Precision and Recall):

	precision	recall	f1-score	support
0	0.82	0.93	0.87	6159
1	0.62	0.35	0.45	1982
accuracy			0.79	8141
macro avg	0.72	0.64	0.66	8141
weighted avg	0.77	0.79	0.77	8141

Area Under the Curve (AUC): 0.640671957669357

(4) Privacy level: $t = 0.2$