650.060 - Small Project in Artificial Intelligence and Cybersecurity

An API for Mutual Information estimation

Francesco Bombassei De Bona Summer Term 2022/2023 12138677

1 Introduction

This project aims to provide a publicly available C++ API for the methodologies introduced in [2] for estimating Mutual Information. A section is dedicated to compare the proposed implementation with the histogram estimator method.

2 Background and terminology

2.1 Mutual Information

Mutual Information (MI) is a measure of the statistical dependence between two random variables. In the context of side-channel analysis, MI is used to quantify the dependence between an observed leakage and a secret key. Higher values of MI indicate a stronger dependence between the two variables. Thus, an higher MI value leads to a higher leakage of information about the secret key by observing a side-channel trace.

The MI between two random variables X and Y is defined as:

$$I(X;Y) = H(X) - H(X|Y) \tag{1}$$
$$= H(X) + H(Y) - H(X,Y) \tag{2}$$

where $H(X)$ is the entropy of X, $H(X|Y)$ is the conditional entropy of X given Y and is defined as

$$H(X|Y) = \begin{cases} \sum_y p(y)H(X|Y=y) & \text{if } Y \text{ is discrete} \\ \int_y p(y)H(X|Y=y)dy & \text{if } Y \text{ is continuous} \end{cases}$$

and $H(X,Y)$ is the joint entropy of X and Y.

2.2 Histogram estimator

The histogram estimator is a method for estimating the MI between two random variables X and Y. This approach is particularly useful when the underlying probability distributions are not known or difficult to model analytically. The core idea behind the histogram estimator is to discretize the continuous random variables into bins and then compute the mutual information based on the joint and marginal frequencies of these bins. The MI between X and Y is then estimated as:

$$I(X;Y) = H(Y) - H(Y|X) \tag{3}$$

where $H(Y)$ is the entropy of Y and $H(Y|X)$ is the conditional entropy of Y given X.

2.3 GKOV estimator

The GKOV estimator is a method for estimating the MI between two random variables X and Y introduced by Gao, Krishnan, Oh and Vishwanath [1]. This estimator can be used on a combination of discrete and continuous random variables. The GKOV estimator is defined as:

$$I_n(X;Y) = \frac{1}{n}\sum_{i=1}^{n}\hat{I}_i = \sum_{i=1}^{n}\left(\psi(\tilde{t}_i) + \log n - \log(n_{x,i}+1) - \log(n_{y,i}+1)\right) \tag{4}$$

where ψ is the digamma function.

3 Implementation

References

[1] Gao, W., Kannan, S., Oh, S., and Viswanath, P. Estimating mutual information for discrete-continuous mixtures, 2017.

[2] Roy, A., Chowdhury, A., and Oswald, E. Leakage Certification based on Consistent MI Estimation, 2022.