State of the Art: Understanding Pointwise Maximal Leakage (PML)

I) Motivations

→ Side-channel : Any unconventional way of getting a slight information about a secret.

→ Mutual information as a leakage measure? Not very adapted here.

II) Maximal Leakage

1) Threat Model and Definition

Setup of a guessing adversary:

- X is a secret
- U is what actually interests the adversary
- Y is the observation of X from which the adversary wants to guess U.

The SSH example

- X are the real keystroke timings
- Y are the observed keystroke timings
- U is the actual password.

We assume we have the Markov chain $U - X - Y$.

Reminder: We say that a Markov graph $G=(V,E)$ holds, if for any separating set S separating V_1 from V_2 (where $V_1 \sqcup S \sqcup V_2 = V$) then

$$V_1 \mid S \perp\!\!\!\perp V_2 \mid S$$

Let $\hat{U}(Y)$ (or even simpler $\hat{U}$) the best estimator of U given Y, i.e. that minimizes $\mathbb{P}(\hat{U}=U)$.

Before observing Y, the probability of guessing U with no clue is $\max_{Z \perp\!\!\!\perp U} \mathbb{P}(Z=U)$ that can be reduced to $\max_{u \in \mathcal{U}} P_U(u)$ where $\mathcal{U}$ is the set of all the possible values taken by U.

The maximal leakage for this specific example is the ratio $\frac{\mathbb{P}(\hat{U}=U)}{\max_{u \in \mathcal{U}} P_U(u)}$.

It means:

"By how much can I multiply my prior probability by observing Y"

It is always preferable to take the log, and $\log_2$ means:

"With a leakage of n bits, I can increase my probability by 2^n"

For a general definition, we would need to fix what the attacker wants. But measuring the leakage of Y over X, this makes no sense that we need to set what is U. In our case, U could be:

- The entire password
- The entire password + more information about the computer
- The first letter of the password
- Whether it is a vowel
- etc...

The SDPI (Strong Processing Data Inequality) allows to maximize over all the possible setups:

Let X and Y be two random variables on respectively $\mathcal{X}$ and $\mathcal{Y}$.

The Maximal Leakage from X to Y is defined by:

$$\mathcal{L}(X \to Y) = \sup_{\substack{U - X - Y - \hat{U} \\ U, \hat{U} \in \mathcal{U}}} \log \frac{\mathbb{P}(\hat{U}=U)}{\max_{u \in \mathcal{U}} P_U(u)}$$

Where $\mathcal{U}$ is any set.

2) Main result

The form provided above is more for the intuition and less for the computation. Indeed, we will show that

$$\mathcal{L}(X \to Y) = \log \sum_{y \in \mathcal{Y}} \max_{\substack{x \in \mathcal{X} \\ P_X(x) > 0}} P_{Y|X}(y|x)$$

This is, by definition, the Sibson mutual information of order infinity, $I_\infty(X;Y)$.
The definition is valid only is $\mathcal{X}$ and $\mathcal{Y}$ are finites alphabets.
If $\mathcal{X} = \{0, 1\}$ then

$$\mathcal{L}(X \to Y) = \log_2\left(1 + \frac{1}{2}\left\|P_{Y|X}(\cdot|0) - P_{Y|X}(\cdot|0)\right\|_1\right)$$

where $\|\cdot\|_1$ is the L_1 distance (Manhattan).

Some properties:

i) If $X - Y - Z$ holds, then $\mathcal{L}(X \to Z) \leq \min\left(\mathcal{L}(X \to Y), \mathcal{L}(Y \to Z)\right)$
↳ Data Processing Inequality

ii) $\mathcal{L}(X \to X) = H_0(X) = \log |\mathrm{Supp}(X)|$

iii) $\mathcal{L}(X \to Y) \leq \min\left(\log|\mathcal{X}|, \log|\mathcal{Y}|\right)$

iv) $\mathcal{L}(X \to Y) \geqslant I(X;Y) \geqslant 0$

v) $\mathcal{L}(X \to Y) = 0 \iff X \perp\!\!\!\perp Y$

vi) If $(X_i, Y_i)_{1 \leqslant i \leqslant l}$ are mutually independent, then

$$\mathcal{L}\left((X_i) \to (Y_i)\right) = \sum_{i=1}^{l} \mathcal{L}(X_i \to Y_i)$$

vii) $\mathcal{L}(X \to Y)$ only depends on $\mathrm{Supp}(X) \subseteq \mathcal{X}$ and $P_{Y|X} \in [0,1]^{\mathcal{Y} \times \mathrm{Supp}(X)}$.

viii) If $\mathrm{Supp}(X)$ is fixed, then $e^{\mathcal{L}(X \to Y)}$ is convex in $P_{Y|X}$.

For a leakage measure, we often consider i), v) and vi) as axiomatic.
Mutual information does check this.

3) Basic notions

It's never a bad idea to get back to the basics.

For a finitely valued random variable X, the entropy of X (or more Shannon's entropy) is defined by:

$$H(X) = \sum_{x \in Supp(X)} p(x) \log\left(\frac{1}{p(x)}\right) = \mathbb{E}\left(\log\left(\frac{1}{p(X)}\right)\right)$$

where p is the distribution of X.

This is the only continuous function $H : \mathcal{X}_{<\infty} \to \mathbb{R}$, that satisfies:

i) $0 \leq H(X) \leq H\left(\frac{1}{n}, \dots, \frac{1}{n}\right)$ where $n = |Supp(X)|$

ii) $H(p_1, \dots, p_n, 0) = H(p_1, \dots, p_n)$

iii) $H(X,Y) = H(X) + H(Y|X)$ this is the Chain Rule

iv) $H\left(\frac{1}{n}, \dots, \frac{1}{n}\right) = \log(n)$ (we can choose any base $b > 1$ we want)

or, equivalently:

i) $\forall \sigma \in S_n$, $H(p_1, \dots, p_n) = H(p_{\sigma(1)}, \dots, p_{\sigma(n)})$

ii) For any $p_n = q_1 + q_2$, $H(p_1, \dots, q_1, q_2) = H(p_1, \dots, p_n) + p_n H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}\right)$

As a measure of uncertainty, we can also mention Rényi entropy for $\alpha \in \mathbb{R}^+ \setminus \{1\}$:

$$H_\alpha(X) = \frac{1}{1-\alpha} \log\left(\sum_{x \in Supp(X)} p(x)^\alpha \right)$$

However, we only have a weak chain rule: "If $X \perp\!\!\!\perp Y$, then $H_\alpha(X,Y) = H_\alpha(X) + H_\alpha(Y)$"

The mutual information measures how much information do two variables share:

$$I(X \to Y) = H(X) + H(Y) - H(X,Y) = H(Y) - H(Y|X)$$

However the fact that it is symmetric is not very meaningful for channels.

We always have that:

i) $0 \leq I(X,Y) \leq H(X), H(Y)$ ii) $I(X,Y)=0 \Leftrightarrow X \perp\!\!\!\perp Y$

This is more a measure of what X and Y have in common than a leakage of the channel $X \to Y$.

Remark: We can try to expand the mutual information:

$$I(X,Y) = H(X) + H(Y) - H(X,Y) = \sum_x \mathbb{P}_X(x) \log\left(\frac{1}{\mathbb{P}_X(x)}\right) + \sum_y \mathbb{P}_Y(y) \log\left(\frac{1}{\mathbb{P}_Y(y)}\right)$$

$$- \sum_{x,y} \mathbb{P}_{X,Y}(x,y) \log\left(\frac{1}{\mathbb{P}_{X,Y}(x,y)}\right) = \sum_{x,y} \mathbb{P}_{X,Y}(x,y) \log\left(\frac{1}{\mathbb{P}_X(x)\mathbb{P}_Y(y)}\right) \sum_{x,y} \mathbb{P}_{X,Y}(x,y) \log\left(\frac{1}{\mathbb{P}_{X,Y}(x,y)}\right)$$

$$\boxed{I(X,Y) = \sum_{x,y} \mathbb{P}_{X,Y}(x,y) \log\left(\frac{\mathbb{P}_{X,Y}(x,y)}{\mathbb{P}_X(x)\mathbb{P}_Y(y)}\right)}$$

A Discrete Memory Channel (DMC) has an input of $\mathcal{X}$ and an output of $\mathcal{Y}$, and induces a transition function $W: \mathcal{Y}|\mathcal{X} \to \mathbb{R}^+$.

In this setup, mutual information can be written as:

$$I(X,Y) = \sum_{x,y} \mathbb{P}_X(x) W(y|x) \log\left(\frac{W(y|x)}{\sum_x \mathbb{P}_X(x) W(y|x)}\right)$$

The capacity is the value of the mutual information, when the input distribution maximizes it: $C = \max_{\mathbb{P}_X} I(X,Y)$.

As we have seen before, the maximal leakage only depends on $\text{Supp}(\mathbb{P}_X)$:

$$\mathcal{L}(X \to Y) = \log \sum_{y \in \mathcal{Y}} \max_{\substack{x \in \mathcal{X} \\ \mathbb{P}_X(x)>0}} W(y|x)$$

What is the relationship between capacity and maximal leakage?

The capacity is a useful tool for the normal framework where a message U is encoded $X^n = f_n(U)$, then passes trough the channel, and is finally decoded : $\hat{U} = g_n(Y^n)$.

Let M_n be an increasing sequence of numbers. We say that $(M_n)_{n\in\mathbb{N}}$ is accepted by the channel if :

$$\sup_{U \in [M_n]} \mathbb{P}(\hat{U} \neq U \mid X^n = f_n(U)) \xrightarrow{n\to+\infty} 0$$

The rate of an accepted $(M_n)_{n\in\mathbb{N}}$, is the value $R = \lim_{n\to+\infty} \frac{1}{n} \log_2(M_n)$ in bits per channel use. The capacity is therefore the supremum over all accepted rates :

$$C = \sup_{(M_n) \text{ accepted}} \lim_{n\to+\infty} \frac{1}{n} \log_2(M_n) \ .$$

If we want to be more formal, we should replace (M_n) by (f_n, g_n, M_n). This is called a "code".

We have the incredible fact that : $C = \max_{P_X \in \Delta(\mathcal{X})} I(X,Y)$ which is constant for a fixed channel. → Discrete...

If we define $C_\alpha = \max_{P_X \in \Delta(\mathcal{X})} I_\alpha(X,Y)$, then we have $\mathcal{L}(X \to Y) = C_\infty$.

Since $I(X,Y) \leq I_\infty(X,Y)$ we therefore have that $\mathcal{L}(X \to Y) \geq C$, always assuming that $\mathrm{Supp}(X) = \mathcal{X}$.

→ For security aspects, Maximal Leakage is more sensitive than Shannon's capacity.

When we define a distance between laws, a good way of measuring the leakage is by measuring the distance between the real joint law and the product of $\mathbb{P}_X$ by another law in $\mathcal{U}$. This way, if $D(\cdot\|\cdot)$ measures the distance between two laws, then:

$$I(X,Y) = \inf_{\mathbb{P}_Y^* \in \mathcal{U}} D(\mathbb{P}_{X,Y} \| \mathbb{P}_X \mathbb{P}_Y^*)$$

If $D(\cdot\|\cdot)$ is the Kullback-Leibler divergence, then $I(X,Y)$ is the mutual information and the symmetry gives us $\mathbb{P}_Y^* = \mathbb{P}_Y = \sum_x \mathbb{P}_{X,Y}(x,\cdot)$.

However, the symmetry does not hold in general. For instance, if we take Rényi divergences, then a good way to define a leakage measure is:

$$I_\alpha(X,Y) = \inf_{\mathbb{P}_Y^* \in \mathcal{U}} D(\mathbb{P}_{X,Y} \| \mathbb{P}_X \mathbb{P}_Y^*)$$

This is the real case. The distance tells us how much information Y gives about X.

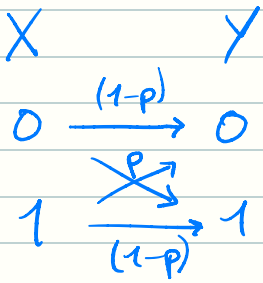
This is the set of all possibilities where Y gives no information about X.

This is not symmetric in general anymore. The maximal leakage is also given by $\mathcal{L}(X \to Y) = I_\infty(X,Y)$ that derives from a distance for laws. Wonderful isn't it?

4) Some examples

Ex 1: $X \sim \mathcal{B}(q)$ $0 < q < 1$ $p \leq \frac{1}{2}$

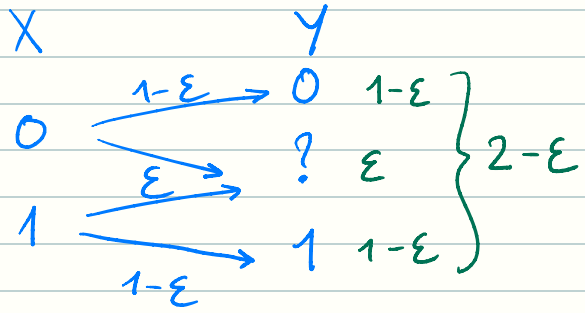
$$\mathcal{L}(X \to Y) = \log(2(1-p)) = 1 + \log(1-p)$$



→ If $p = 0$, I can increase the guessing probability by $2^1 = 2$.

→ If $p = \frac{1}{2}$, I cannot increase my guessing probability.

Ex 2: $X \sim B(q)$ $0<q<1$ $0 \leq \varepsilon < 1$



$$\mathcal{L}(X \to Y) = \log(2-\varepsilon)$$

Remark: i) For any deterministic channel, $\mathcal{L}(X \to Y) = \log(|Supp(Y)|)$. We also have that $|Supp(Y)| \leq |Supp(X)|$.

ii) In the definition of maximal leakage, a more intuitive version could be:

$$\mathcal{L}(X \to Y) = \sup_{\substack{u=f(x) \\ \hat{u}=f(y)}} \log \frac{\mathbb{P}(u=\hat{u})}{\max_{u \in \mathcal{U}} P_U(u)}$$

5) Proof of the main result

We want to prove the following:

$$\sup_{\substack{u-x-y-\hat{u} \\ u,\hat{u} \in \mathcal{U}}} \log \frac{\mathbb{P}(\hat{u}=u)}{\max_{u \in \mathcal{U}} P_U(u)} = \log \sum_{y \in \mathcal{Y}} \max_{\substack{x \in \mathcal{X} \\ P_X(x)>0}} P_{Y|X}(y|x)$$

Proof: Let $\mathcal{L}(X \to Y)[U] = \log \frac{\sum_y \max_{u \in \mathcal{U}} P_{UY}(u,y)}{\max_{u \in \mathcal{U}} P_U(u)}$. This is clearly the best estimator we can get, and therefore $\mathcal{L}(X \to Y) = \sup_{U-X-Y} \mathcal{L}(X \to Y)[U]$.

The probability that the best estimator actually gives the best result can be developed by the Markov chain $U-X-Y$:

$$\sum_y \max_{u \in \mathcal{U}} P_{UY}(u,y) = \sum_y \max_{u \in \mathcal{U}} \sum_x P_X(x) P_{U|X}(u,x) P_{Y|X}(y,x)$$

$$\leq \sum_y \max_{u \in \mathcal{U}} \sum_x P_X(x) P_{U|X}(u,x) \left[\max_{x' \in \mathcal{X}} P_{Y|X}(y,x')\right]$$

$$= \sum_y \left[\max_{x' \in \mathcal{X}} P_{Y|X}(y,x')\right] \max_{u \in \mathcal{U}} \sum_x P_X(x) P_{U|X}(u,x)$$

$$= \max_{u \in \mathcal{U}} P_U(u) \sum_y \left[\max_{x \in \mathcal{X}} P_{Y|X}(y,x)\right]$$

So LHS $\leq$ RHS. For the reversed inequality, we construct a $P_{U|X}$ that reaches the RHS.

Let $p^* = \min_{x \in \mathcal{X}} P_X(x) > 0$ (we can assume $\text{Supp}(X) = \mathcal{X}$ without any loss of generality). For each $x \in \mathcal{X}$ let $k(x) = \frac{P_X(x)}{p^*}$.

We first define $\mathcal{U} = \mathcal{X} \times [\![1, \lceil k(x) \rceil]\!]$, or more formally, $\mathcal{U}$ is defined by $\bigcup_{x \in \mathcal{X}} \{(x,1), \dots, (x, \lceil k(x) \rceil)\}$.

For $x \in \mathcal{X}$, U returns an element of $\{(x,1), \dots, (x, \lfloor k(x) \rfloor)\}$ with the same probability equal to $\frac{p^*}{P_X(x)}$ and returns $(x, \lceil k(x) \rceil)$ with the resting probability $1 - \lfloor k(x) \rfloor \frac{p^*}{P_X(x)}$.

We have $\max_{u \in \mathcal{U}} P_U(u) = p^*$ obtained for the value $x \in \mathcal{X}$ such that $P_X(x) = p^*$.
Moreover, $\sum_y \max_{u \in \mathcal{U}} P_{UY}(u,y) = \sum_y \max_{u \in \mathcal{U}} \sum_x P_X(x) P_{U|X}(u|x) P_{Y|X}(y|x)$

$$= \sum_y \max_{x \in \mathcal{X}} P_X(x) P_{U|X}((x,1)|x) P_{Y|X}(y|x)$$

$$= p^* \sum_y \max_{x \in \mathcal{X}} P_{Y|X}(y|x)$$

We just proved that LHS $\geq$ RHS.