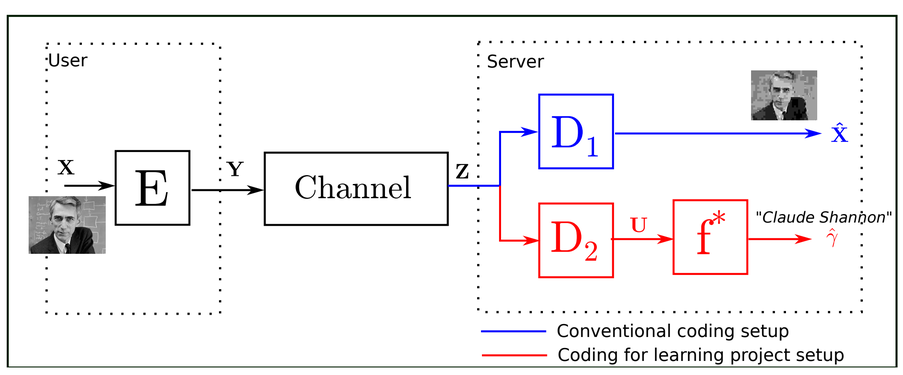
CoLearn

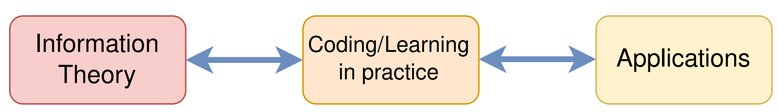
Coding for Learning

1. Project scientific objectives

Context: Huge mass of data (images, video, etc.) need to be sorted, processed, stored, recommended to users, etc.



- **Objective**: **Learning and data reconstruction over coded data**
- Key questions:
  - Is there a tradeoff between the data reconstruction and learning objectives?
  - Can one perform learning without prior decoding?
  - Does the source-channel separation principle still hold?
- Approach:

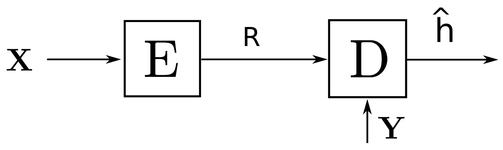


2. Information-Theoretic bounds for Regression

Problem addressed:

- Few is known about IT limits of communication-for-learning schemes
- We consider **regression** as a first yet simple learning problem

$$X = \sum_{k=1}^{K} \alpha_k h_k(Y) + \epsilon$$



Training sequence $(\mathbf{X}, \mathbf{Y})$, Test sequence $(\tilde{\mathbf{X}}, \tilde{\mathbf{Y}})$

Minimum expected loss: $L^\star = \inf_f E\left[(X - f(Y))^2\right]$

Expected Generalization error (GE): $G^{(k)}(\widehat{f}) = E_{\mathbf{X},\mathbf{Y}}\left[E\left[(\tilde{X} - \hat{f}(\tilde{Y}))^2 | \mathbf{X}, \mathbf{Y}\right]^k\right]$

Asymptotic Rate-GE region

- Existing bounds by **Raginsky**, for a given R:

$$(L^\star)^{1/2} \leq \limsup_{n\to\infty} G^{(1/2)}(\hat{f}) \leq (L^\star)^{1/2} + 2\mathbb{D}_{X|Y}(R)^{1/2}$$

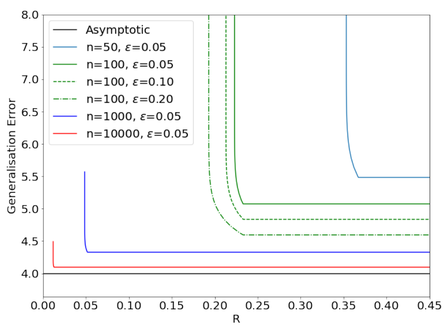
- We proposed an **IT scheme** which achieves the min. for any $R > 0$:

$$\limsup_{n\to\infty} G^{(1)}(\hat{f}) = L^\star$$

Finite-Length Rate-GE region

- **Excess probability**: $\epsilon = \mathbb{P}(G^{(1)}(\widehat{f}) \geq g)$
- **Loss-information density vector**:

$$\mathbf{i}(U, X, Y, \tilde{X}, \tilde{Y}) := \begin{bmatrix} -\log \dfrac{P_{U|Y}(U|Y)}{P_U(U)} \\ \log \dfrac{P_{U|X}(U|X)}{P_U(U)} \\ \ell(\tilde{X}, \hat{f}^{(n)}(\mathbf{Z}, \tilde{Y})) \end{bmatrix}$$



- **Rate-GE region**:

$$R(n, \varepsilon, g) \leq \inf\left\{\mathbf{M}\left(\mathbb{E}[\mathbf{i}] + \frac{\mathcal{S}(\mathbb{C}[\mathbf{i}], \varepsilon)}{\sqrt{n}} + \frac{2\log n}{n}\mathbf{I}_3\right)\right\}$$

We showed that there is **no tradeoff** between Distortion and GE

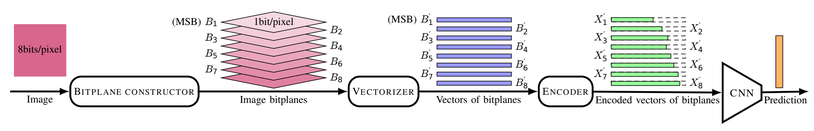
Perspectives:

- Extend the results to non-parametric regression, to other metrics
- Practical coding scheme for regression, using non-binary LDPC codes
- Consider other learning problems, such as classification

3. Classification from Entropy-Coded Images

Problem addressed:

- Entropy-coding breaks the data structure
- Can we do learning without any prior decoding? **No? Let's try!**
- We consider **image classification** as a first yet simple learning problem



Results on entropy-coded grayscales images :

- We considered CNN architectures designed for 1D data

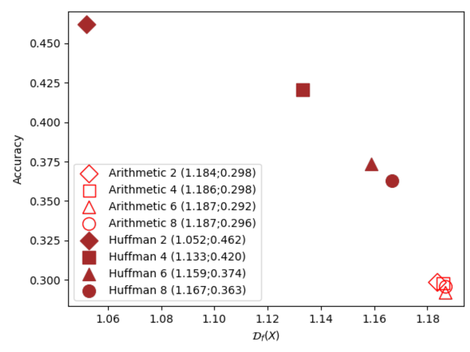
	Coding Type			
Dataset	None	Huffman	Arithmetic	JPEG
MNIST	0.98911	0.83234	0.63130	-
Fashion-MNIST	0.90189	0.76347	0.68987	-
YCIFAR-10	**0.56573**	**0.36062**	**0.29762**	**0.32459**

Predicting accuracy loss:

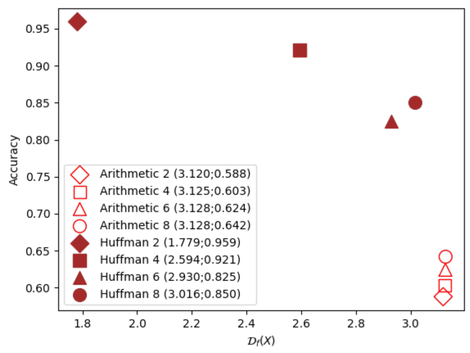
- **Proposed Metric**:

$$\mathcal{D}_f(X) = \frac{\sum_{i=0}^{B} \text{ApEn}_{m,r}(f(B_i(X)))}{\sum_{i=0}^{B} \text{ApEn}_{m,r}(B_i(X))}$$

- $\text{ApEn}_{m,r}$: approximate entropy with window size m and threshold r
- f entropy coding function, B_i functional extractor of bitplane i



On YCIFAR-10



On MNIST

Perspectives:

- Can we learn decoding?
- How to order computational cost of accessing a given data in the coded bitstream?

4. Perspectives

- Classify learning applications depending on whether there is a tradeoff between data reconstruction and learning
- Develop practical coding schemes addressing both data reconstruction and learning
- Consider more complex learning problems and communication conditions related to the project applications (video coding, submarine communications)

- Rémi Piau, Thomas Maugey, Aline Roumy, Predicting CNN learning accuracy using chaos measurement, ICASSP 2023
- Jiahui Wei, Elsa Dupraz, Philippe Mary, Régions atteignables pour la régression linéaire sur données compressées avec information adjacente, GRETSI 2023
- Rémi Piau, Thomas Maugey, Aline Roumy, Prédiction de la précision d'apprentissage des réseaux de neurones convolutifs par mesure du chaos, GRETSI 2023
- Jiahui Wei, Elsa Dupraz, Philippe Mary, Asymptotic and non-asymptotic rate-loss bounds for linear regression with side information, EUSIPCO 2023
- Remi Piau, Aline Roumy, and Thomas Maugey, Learning on entropy coded images with CNN, ICASSP 2023
- Alireza Tasdighi and Elsa Dupraz, An End-to-End Scheme for Learning over Compressed Data Transmitted Through a Noisy Channel, IEEE Access, 2023