Learning over Compressed Data

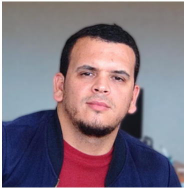
Elsa Dupraz

June 2025

Coding for Learning team

Ismaila
Salihou-Adamou

Jiahui Wei



Ahcen Aliouat

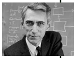
Yann Miguet

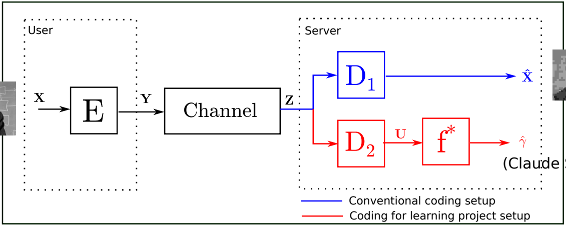
Motivation

- Every minute:
  - 500 hours of video uploaded on Youtube
  - $240,000$ images uploaded on Facebook
- Huge mass of data for transmission, storage, and processing
- **Need for learning over coded data**

Motivation

- Every minute:
  - 500 hours of video uploaded on Youtube
  - $240,000$ images uploaded on Facebook
- Huge mass of data for transmission, storage, and processing
- **Need for learning over coded data**
- Goal-oriented communications/task-aware compression:

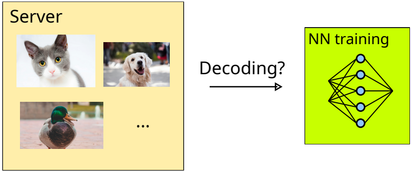




(Claude Shannon)

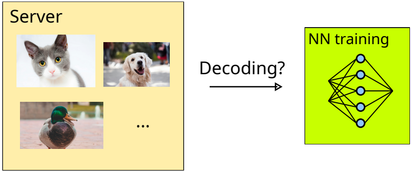
(Some) key questions

- Can we apply learning over compressed data, **without any decoding**?

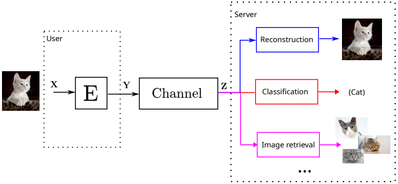


(Some) key questions

- Can we apply learning over compressed data, **without any decoding**?

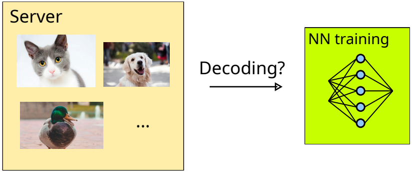


- Can we design a **universal** coding scheme?

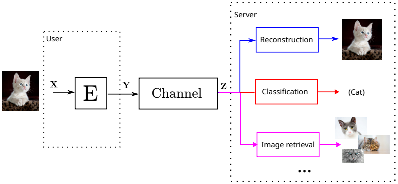


(Some) key questions

- Can we apply learning over compressed data, **without any decoding**?



- Can we design a **universal** coding scheme?



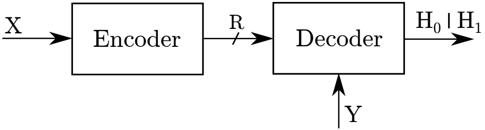
In this talk

- What is the **IT performance** of coding schemes for learning?
- Can we use the IT analysis to design efficient **practical coding schemes**?

Outline

Distributed Hypothesis Testing[2]

1



- Hypothesis-testing formulation:

$$\mathcal{H}_0 : (\mathbf{X}, \mathbf{Y}) \sim P_{\mathbf{XY}},$$
$$\mathcal{H}_1 : (\mathbf{X}, \mathbf{Y}) \sim P_{\bar{\mathbf{X}}\bar{\mathbf{Y}}}.$$

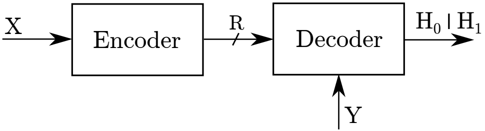
- Performance criteria:

$$\text{(Type-I error)} \quad \alpha_n = \mathbb{P}\left(\text{decide } \mathcal{H}_1 \mid \mathcal{H}_0 \text{ is true}\right),$$
$$\text{(Type-II error)} \quad \beta_n = \mathbb{P}\left(\text{decide } \mathcal{H}_0 \mid \mathcal{H}_1 \text{ is true}\right).$$

[1] R. Ahlswede and I. Csiszár, Hypothesis testing with communication constraints," IEEE Transactions on Information Theory, 1986

[2] T. S. Han, Hypothesis Testing with Multiterminal Data Compression, IEEE Transactions on Information Theory, 1987.

Information-theoretic results

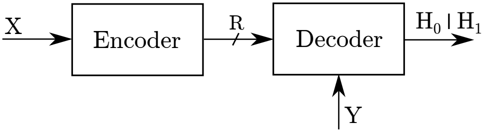


- **Objective**: for a given R, derive the **error exponent** θ such that

$$\alpha_n < \alpha, \quad \lim_{n\to\infty} \sup \frac{1}{n} \log \frac{1}{\beta_n} \geq \theta \quad (\beta_n \leq \exp(-n\theta))$$

[3] Shimokawa et al. (1994), Katz et al. (2015), Kochman et al. (2023), Adamou et al. (2024), etc.

Information-theoretic results



- **Objective**: for a given R, derive the **error exponent** θ such that

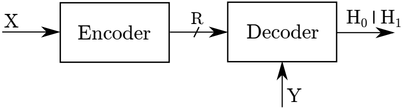
$$\alpha_n < \alpha, \quad \lim_{n\to\infty} \sup \frac{1}{n} \log \frac{1}{\beta_n} \geq \theta \quad (\beta_n \leq \exp(-n\theta))$$

- It was shown that [3]

$$\theta \leq \sup_{P_{\mathbf{U}|\mathbf{X}}} \left\{ \min \left\{ \theta_{test}, \theta_{bin} \right\} \right\},$$

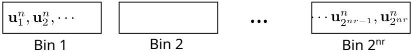
[3] Shimokawa et al. (1994), Katz et al. (2015), Kochman et al. (2023), Adamou et al. (2024), etc.

Information-theoretic coding scheme

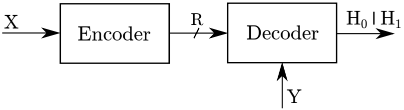


Encoder:

- **Quantization**: represent the 2^n sequences $\mathbf{x}^n$ by 2^{nr} codewords $\mathbf{u}^n$
- **Binning**: randomly assign the 2^{nr} codewords to 2^{nR} bins ($R < r$)

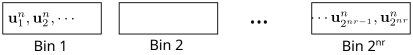


Information-theoretic coding scheme



Encoder:

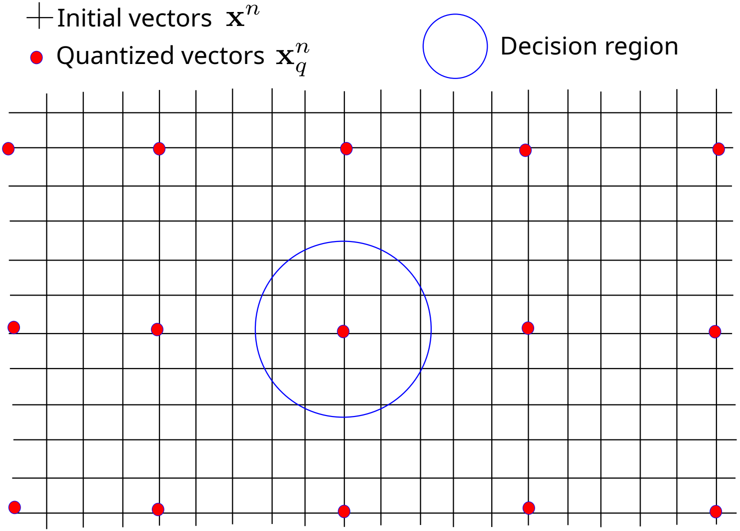
- **Quantization**: represent the 2^n sequences $\mathbf{x}^n$ by 2^{nr} codewords $\mathbf{u}^n$
- **Binning**: randomly assign the 2^{nr} codewords to 2^{nR} bins $(R < r)$



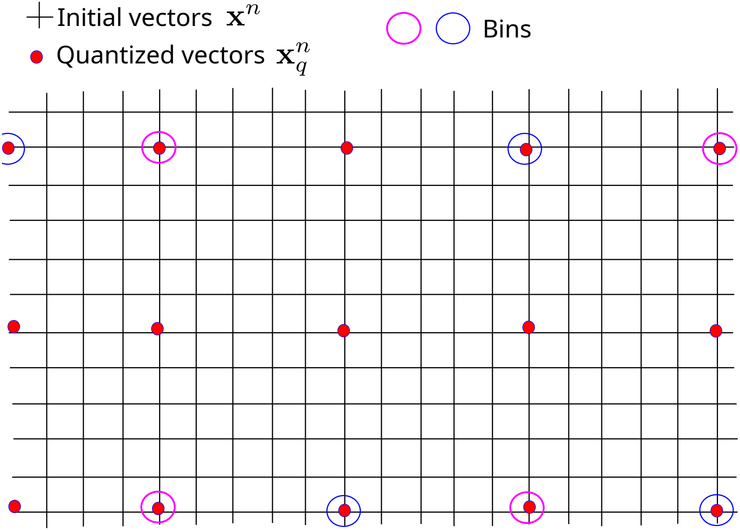
Decoder:

- **Bin extraction**: Pick $\hat{\mathbf{u}}^n$ such that $(\hat{\mathbf{u}}^n, \mathbf{y}^n) \in T_n^{(2)}$
- **Decision**: Decide $\mathcal{H}_0$ if $(\hat{\mathbf{u}}^n, \mathbf{y}^n) \in A_n$

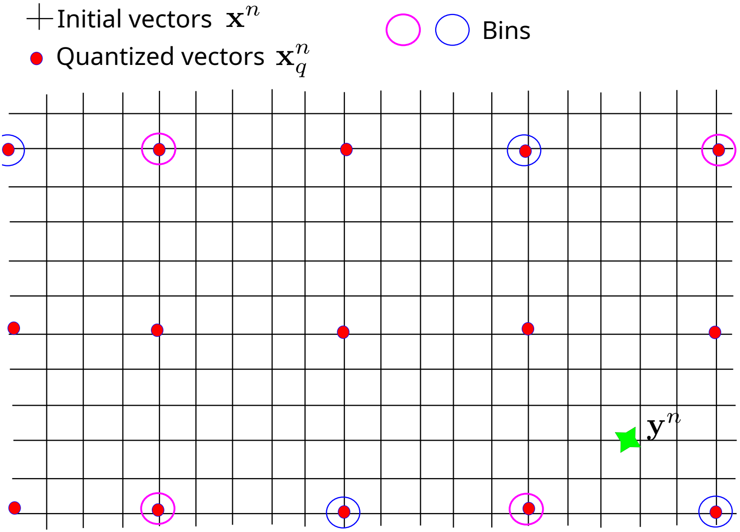
Quantize-binning scheme in practice



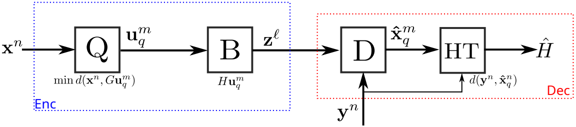
Quantize-binning scheme in practice



Quantize-binning scheme in practice



Proposed coding scheme[4]



Encoder: (G_q: $n \times k$ and H_b: $\ell \times k$)

- **Quantization**: Calculate $\mathbf{u}_q^k$ as:

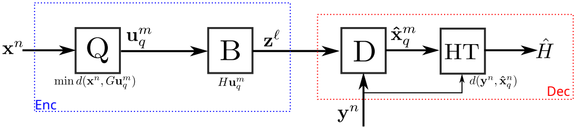
$$\mathbf{u}_q^k = \arg\min_{\mathbf{u}^k} d(G_q \mathbf{u}^k, \mathbf{x}^n) \quad \text{with } \mathbf{x}_q^n = G_q \mathbf{u}^k$$

- **Binning**: we compute

$$\mathbf{z}^\ell = H_b \mathbf{u}_q^k$$

[4] Elsa Dupraz, Ismaila Salihou Adamou, Reza Asvadi, Tadashi Matsumoto, Practical Short-Length Coding Schemes for Binary Distributed Hypothesis Testing, ISIT 2024

Proposed coding scheme[4]



Encoder: (G_q: $n \times k$ and H_b: $\ell \times k$)

- **Quantization**: Calculate $\mathbf{u}_q^k$ as:

$$\mathbf{u}_q^k = \arg\min_{\mathbf{u}^k} d(G_q \mathbf{u}^k, \mathbf{x}^n) \quad \text{with } \mathbf{x}_q^n = G_q \mathbf{u}^k$$

- **Binning**: we compute

$$\mathbf{z}^\ell = H_b \mathbf{u}_q^k$$

Decoder:

- **Bin extraction**: Identify the vector $\mathbf{u}_q^k$ such that

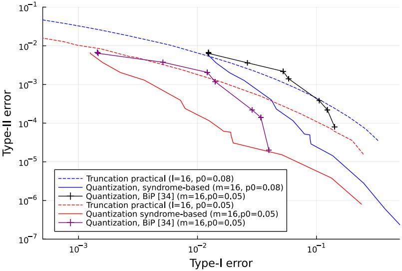
$$\hat{\mathbf{u}}_q^k = \arg\min_{\mathbf{u}_q^k} d(G_q \mathbf{u}_q^k, \mathbf{y}^n) \text{ s.t. } H_b \mathbf{u}^k = \mathbf{z}^\ell$$

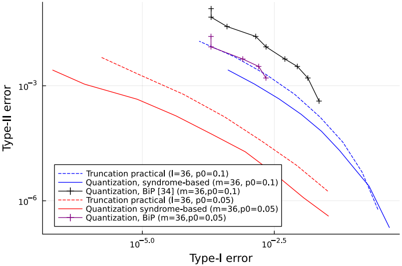
- **Decision**: Apply a NP test as $d(\mathbf{x}_q^n, \mathbf{y}^n) < \lambda$ to decide $\mathcal{H}_0$

[4] Elsa Dupraz, Ismaila Salihou Adamou, Reza Asvadi, Tadashi Matsumoto, Practical Short-Length Coding Schemes for Binary Distributed Hypothesis Testing, ISIT 2024

Numerical results

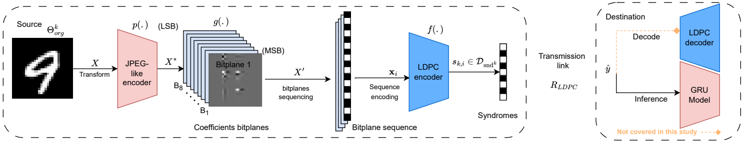
- We need to consider short codes
- Practical quantizers with BCH(31,16) and BCH(63,36) codes





Learning over LDPC-coded data[5]

- Classification over compressed images, without any prior decoding



- Accuracy of the GRU model on LDPC-coded data:

Dataset	Model	No coding		Coding on Orig. (Setup1)			Coding on JPEG (Setup2)				
		None	None MSB	Huff [28]	Arith [28]	LDPC	JPEG [28]	DCT -tr. [27]	J-L 8bp	J-L MSB	J-L MSB+1bp
MNIST	**GRU12(proposed)**	0.9439	0.8842	0.6790	0.5086	**0.8192**	-	-	**0.9060**	0.6548	0.8791
	GRU32(proposed)	0.9799	0.9154	0.7563	0.5370	**0.8556**	-	-	**0.9237**	0.6843	0.8849
	UVGG11 [28]	0.9891	-	0.8323	0.6313	-	-	-	-	-	-
	URESNET18 [28]	0.9875	-	0.7450	0.5949	-	-	-	-	-	-
	FullyConn [27]	0.9200	-	-	-	-	-	0.9000	-	-	-
YCIFAR -10	**GRU12**	0.3127	0.3249	0.2374	-	**0.4070**	-	-	**0.4234**	0.1350	0.3537
	GRU32	0.3596	0.3560	0.2400	-	**0.4171**	-	-	**0.4316**	0.1403	0.3544
	UVGG11 [28]	0.5657	-	0.3606	0.2976	-	0.3245	-	-	-	-
	URESNET18 [28]	0.3836	-	0.2591	0.2432	-	-	-	-	-	-
	FullyConn [27]	0.3800	-	-	-	-	-	0.3000	-	-	-

[5] A. Aliouat, E. Dupraz, Learning on JPEG-LDPC Compressed Images: Classifying with Syndromes, EUSIPCO 2024

Outline

Conclusion

- Various tasks can be addressed with **quantize-binning schemes** constructed using linear block codes (regression, classification, clustering)
- There is a need for more systematic **information-theoretic analyses** of these tasks
- Issues of **learning without decoding** and **universal codes** are still widely open

Conclusion

- Various tasks can be addressed with **quantize-binning schemes** constructed using linear block codes (regression, classification, clustering)
- There is a need for more systematic **information-theoretic analyses** of these tasks
- Issues of **learning without decoding** and **universal codes** are still widely open

Thank you!