Deep Kernel Representation Learning for Complex Data and Reliability Issues

Pierre LAFORGUE, LTCI, Télécom Paris, France

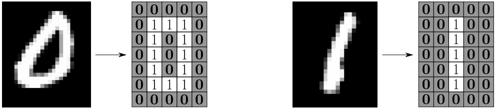
Supervisors: Florence d'Alché-Buc, Stephan Clémençon

Motivation: need for structured data representations

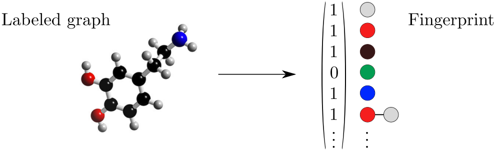
Goal of ML: infer from a set of examples, the relationship between some explanatory variables x, and a target output y

A representation: set of features characterizing the observations

Ex 1:
digit recognition (MNIST)



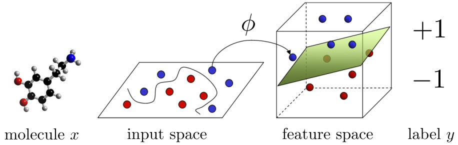
Ex 2:
molecule activity prediction



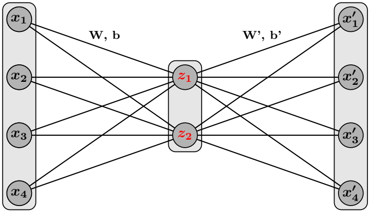
How to (automatically) learn structured data representations?

The Kernel Autoencoder: building blocks

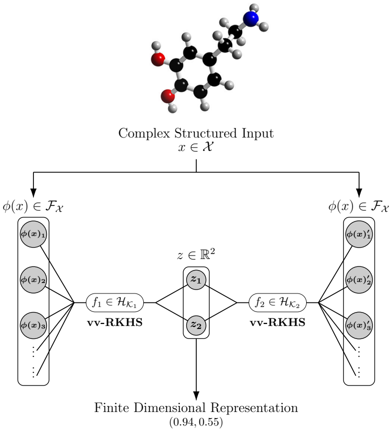
How to deal with non-vectorial data in ML? **Kernel Methods**



How to learn representations of vectorial data in ML? **Autoencoders**



The Kernel Autoencoder [Laforgue et al., 2019]



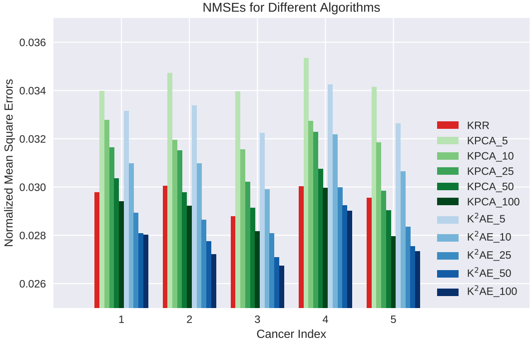
$$\textbf{K}^2\textbf{AE:} \min_{f_l \in \text{vv-RKHS}} \frac{1}{n} \sum_{i=1}^{n} \left\| \phi(x_i) - f_L \circ \ldots \circ f_1(\phi(x_i)) \right\|_{\mathcal{F}_\mathcal{X}}^2 + \sum_{l=1}^{L} \lambda_l \|f_l\|_{\mathcal{H}_l}^2$$

The Kernel Autoencoder: results

On the theoretical side:

- Connection to Kernel PCA [Schölkopf et al., 1997]
- Generalization guarantees through vectorial Rademacher complexities
- Representer Theorem and optimization procedure

On the practical side:

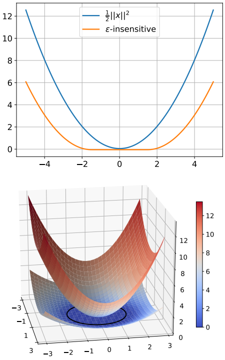


Robust losses in vv-RKHS: motivations

Kernel Autoencoder.

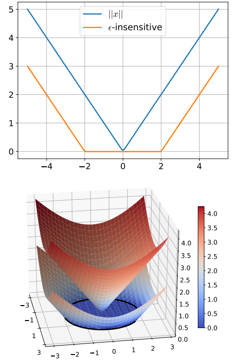
$$\min_{h_1, h_2 \in \mathcal{H}^1_{\mathcal{K}} \times \mathcal{H}^2_{\mathcal{K}}} \quad \frac{1}{2n} \sum_{i=1}^{n} \left\| \phi(x_i) - h_2 \circ h_1(\phi(x_i)) \right\|^2_{\mathcal{F}_{\mathcal{X}}} + \Lambda \operatorname{Reg}(h_1, h_2)$$

ϵ-Ridge



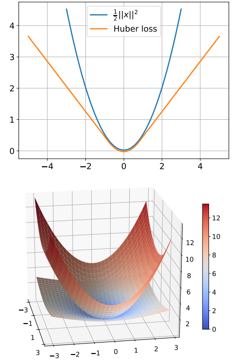
(Sparsity)

ϵ-SVR



(Sparsity, Robustness)

κ-Huber



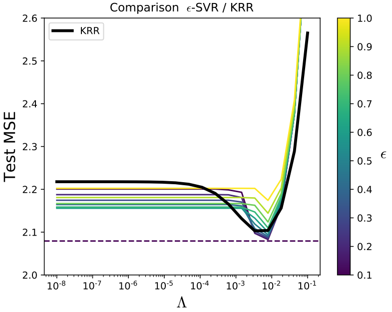
(Robustness)

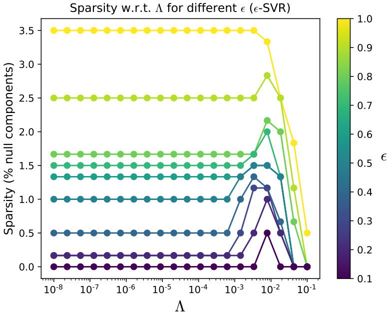
Robust losses in vv-RKHS: results

On the theoretical side:

- Double Representer Theorem: coeffs are linear comb. of the outputs
- The dual optimization problems are well known
- Algorithmic stability analysis

On the practical side:





Conclusion

1. The Kernel Autoencoder allows to extract vectorial representation from structured data
2. Using more complex loss functions is possible and can bring robustness
3. Robustness and reliability can also be achieved by MoM-ifying or debiasing the ERM criterion

Thanks to: Florence d'Alché-Buc, Stephan Clémençon, Alex Lambert, Luc Brogat-Motte, Kevni Massias