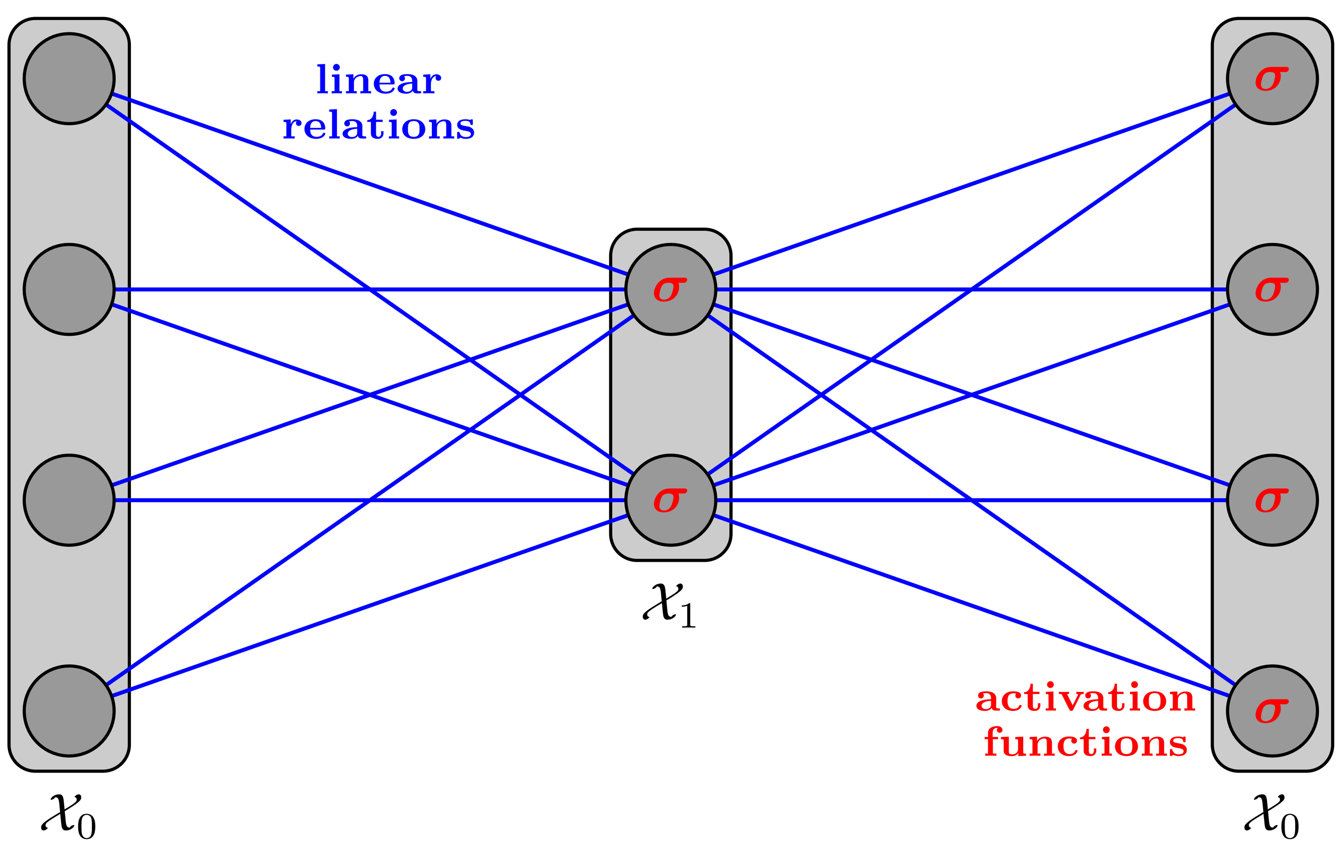
# Autoencoding any Data through Kernel Autoencoders

P. Laforgue⋆, S. Clémençon⋆, F. d'Alché-Buc⋆

⋆ LTCI, Télécom ParisTech, Université Paris-Saclay, 75013, Paris, France

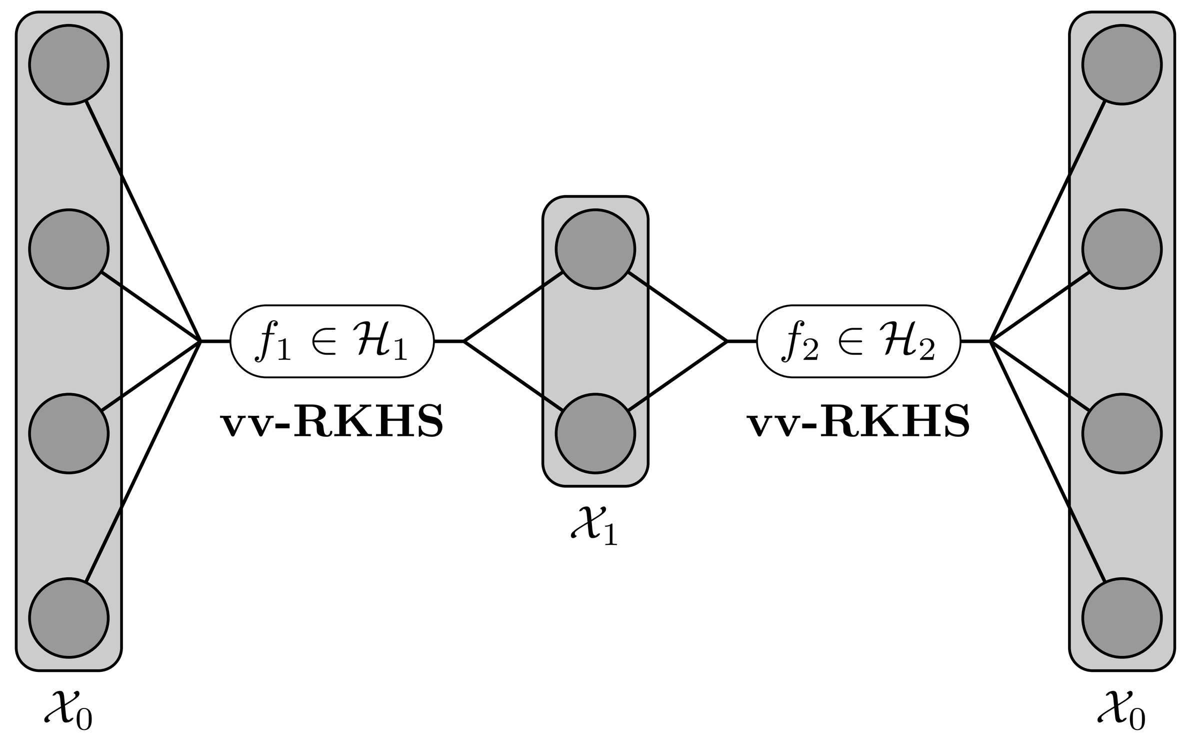
## Standard Autoencoder (AE)

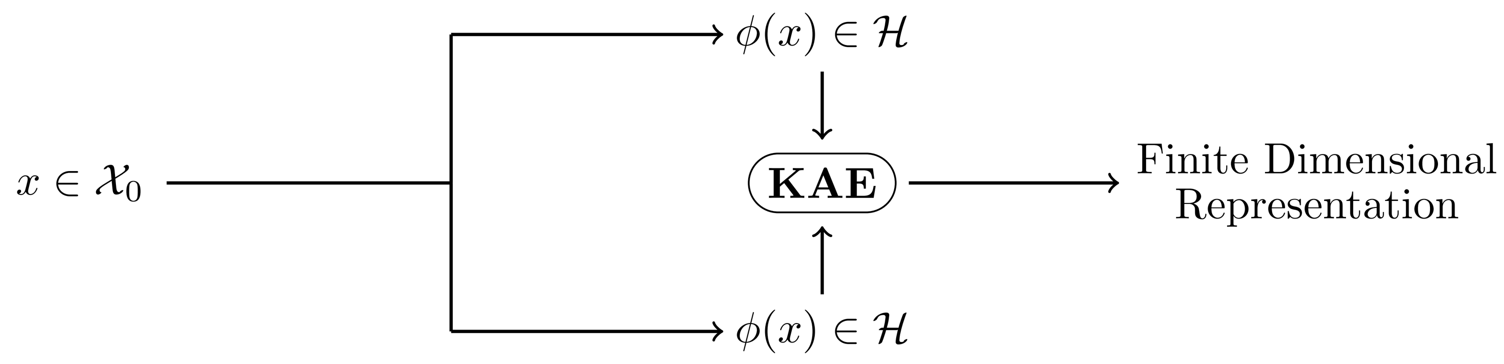
- $\mathcal{S} = (x_1, \ldots, x_n)$ i.i.d. sample on $\mathcal{X}_0 = \mathbb{R}^d$. $\mathcal{X}_1 = \mathbb{R}^p$ with $p < d$
- $f : \mathcal{X}_0 \to \mathcal{X}_1, \quad f(x) = \sigma(W_1 x + b_1), \quad W_1 \in \mathbb{R}^{p \times d}, b_1 \in \mathbb{R}^p$
- $g : \mathcal{X}_1 \to \mathcal{X}_0, \quad g(y) = \sigma(W_2 y + b_2), \quad W_2 \in \mathbb{R}^{d \times p}, b_2 \in \mathbb{R}^d$
- $\min_{W_1, W_2, b_1, b_2} \quad \frac{1}{n} \sum_{i=1}^n \|x_i - g \circ f(x_i)\|^2_{\mathcal{X}_0}$




## Kernel Autoencoder (KAE)

- $\mathcal{X}_0, \mathcal{X}_1$ Hilbert spaces endowed with OVKs $\mathcal{K}_1 : \mathcal{X}_0 \times \mathcal{X}_0 \to \mathcal{L}(\mathcal{X}_1)$ and $\mathcal{K}_2 : \mathcal{X}_1 \times \mathcal{X}_1 \to \mathcal{L}(\mathcal{X}_0)$, associated to Vector Valued RKHSs $\mathcal{H}_1$ and $\mathcal{H}_2$
- $\min_{f_1 \in \mathcal{H}_1,\ f_2 \in \mathcal{H}_2} \quad \frac{1}{n} \sum_{i=1}^n \|x_i - f_2 \circ f_1(x_i)\|^2_{\mathcal{X}_0} + \lambda_1 \|f_1\|^2_{\mathcal{H}_1} + \lambda_2 \|f_2\|^2_{\mathcal{H}_2}$




## Representer Theorem

Let $L \in \mathbb{N}$, and $V : \mathcal{X}_L^n \times \mathbb{R}_+^L \to \mathbb{R}$ a function of $n + L$ variables, strictly increasing in each of its $L$ last arguments. Suppose that $(f_1^*, \ldots, f_L^*)$ is a minimizer on $\mathcal{H}_1 \times \ldots \times \mathcal{H}_L$ of:

$$V\Big((f_L \circ \ldots \circ f_1)(x_1), \ldots, (f_L \circ \ldots \circ f_1)(x_n), \|f_1\|_{\mathcal{H}_1}, \ldots, \|f_L\|_{\mathcal{H}_L}\Big)$$

Let $x_i^{*(l)} := f_l^* \circ \ldots \circ f_1^*(x_i)$, $x_i^{*(0)} := x_i$. Then, $\exists\, (\varphi_{1,1}^*, \ldots, \varphi_{L,n}^*) \in \mathcal{X}_1^n \times \ldots \times \mathcal{X}_L^n$ such that:

$$\forall\, l \leq L, \quad f_l^*(\cdot) = \sum_{i=1}^n \mathcal{K}_l\Big(\ \cdot\ , x_i^{*(l-1)}\Big)\, \varphi_{l,i}^*$$

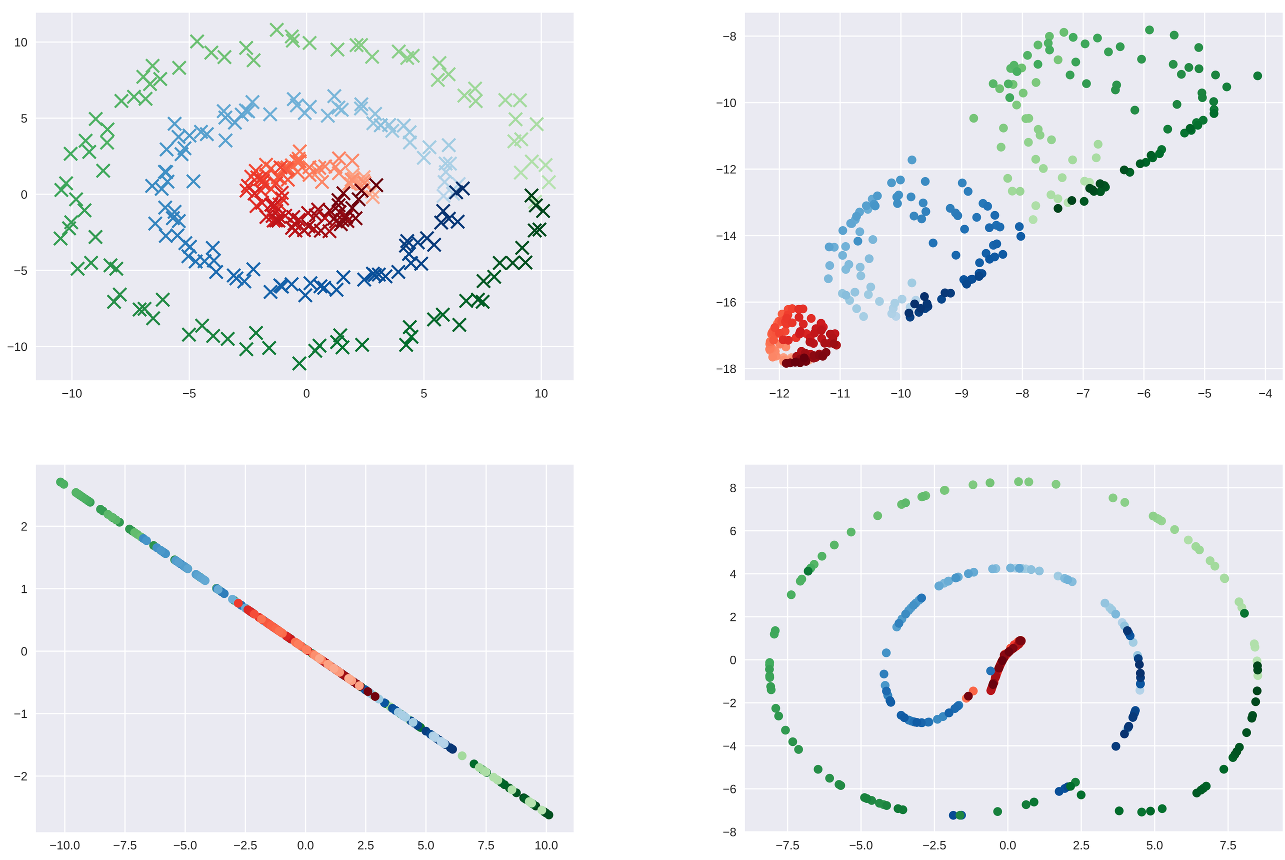
## Algorithm

**Algorithm 1** General Hilbert KAE and $K^2$AE

**input** : Gram matrix $K_{in}$
**init** : $\Phi_1 = \Phi_1^{init}, \ldots, \Phi_{L-1} = \Phi_{L-1}^{init}$,
$N_L = N_{\text{KRR}}(\Phi_1, \ldots, \Phi_{L-1}, K_{in}, \lambda_L)$
**for** *epoch t from* 1 *to* $T$ **do**
  `// inner coefficients updates at fixed` $N_L$
  **for** *layer l from* 1 *to* $L-1$ **do**
    $\Phi_l = \Phi_l - \gamma_t\, \nabla_{\Phi_l}(\hat{\epsilon}_n + \Omega \mid N_L)$
  `//` $N_L$ `update`
  $N_L = N_{\text{KRR}}(\Phi_1, \ldots, \Phi_{L-1}, K_{in}, \lambda_L)$
**return** $\Phi_1, \ldots, \Phi_{L-1}$

## Unsupervised Experiments

**1)** Original data, 2D hidden layer, recontructions by 2-1-2 standard and kernel AEs.



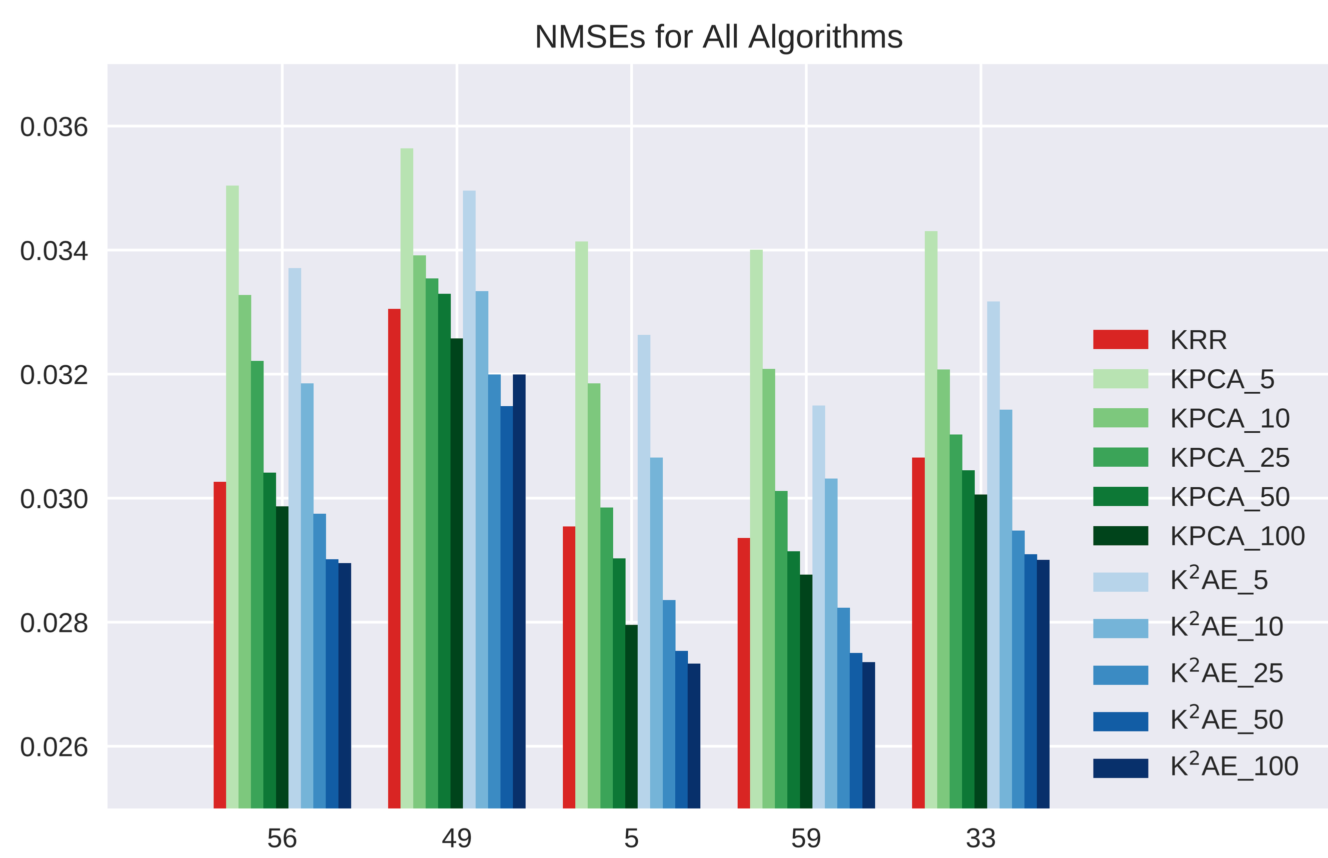
**2)** MSREs on Test Metabolites

| Dimension | AE (sigmoid) | AE (relu) | KAE |
|---|---|---|---|
| 5 | 99.81 | 96.62 | **76.38** |
| 10 | 87.36 | 84.02 | **65.76** |
| 25 | 72.31 | 68.77 | **51.63** |
| 50 | 63.00 | 58.29 | **40.72** |
| 100 | 55.43 | 48.63 | **36.27** |

## Supervised Experiments

$K^2$AE run on molecules after a Tanimoto kernel transformation. Then Random Forests are fed with these finite-dimensional representations. The table stores the Normalized Mean Squared Errors (NMSEs):

| | KRR | $\text{KPCA}_{10}$+RF | $\text{KPCA}_{50}$+RF | $K^2\text{AE}_{10}$+RF | $K^2\text{AE}_{50}$+RF |
|---|---|---|---|---|---|
| C.1 | 0.0298 | 0.0328 | 0.0304 | 0.0310 | **0.0281** |
| C.2 | 0.0300 | 0.0319 | 0.0298 | 0.0310 | **0.0278** |
| C.3 | 0.0288 | 0.0316 | 0.0291 | 0.0299 | **0.0271** |




## References

C. Micchelli, and M. Pontil. On learning vector-valued functions. Neural computation, 17(1): 177–204, 2005

P. Baldi. Autoencoders, Unsupervised Learning, and Deep Architectures. Proc. of JMLR, Workshop *Unsupervised and Transfer Learning*, 2012