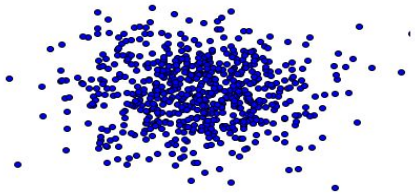
Stochastic Deep Networks

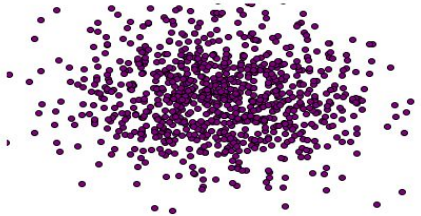
Gwendoline De Bie, Gabriel Peyré, Marco Cuturi

Deep Architectures on Density Inputs

t=1, *1000 cells*



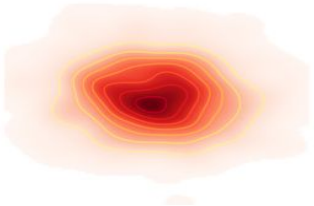
t=2, *650 cells*

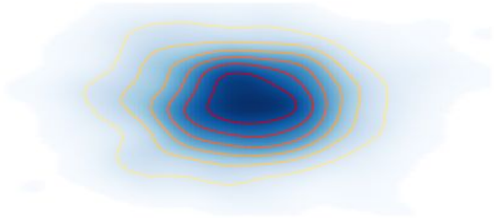


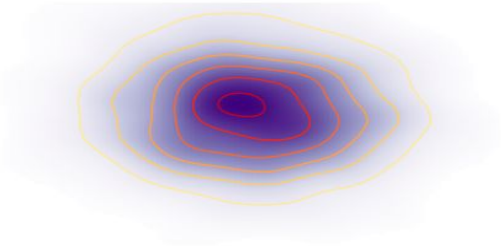
t=3, *890 cells*

of varying physical attributes

➢ Representing inputs as **densities** (discretized in practice)

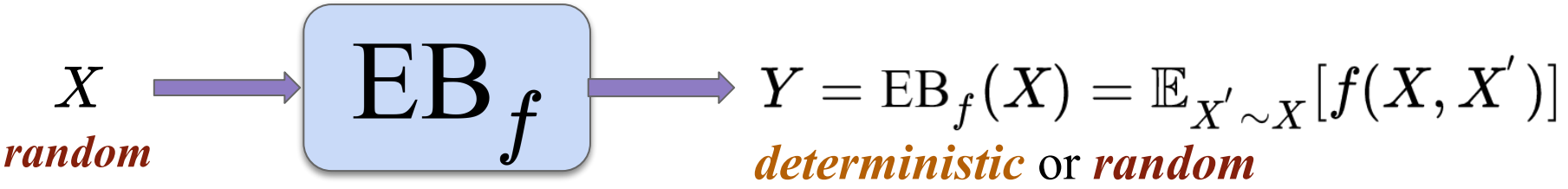






➢ How to define a **‘layer’** of a Deep Net taking such inputs ?

Proposed Layer: Elementary Block (EB)



Discrete case: $X \sim \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i}$

$\rightarrow Y \sim \frac{1}{n}\sum_{i=1}^{n} \delta_{y_i}$, where: $y_i = \sum_{j=1}^{n} f(x_i, x_j)$

Fully connected case: $f(x,u) = (\phi(y_i))_i$

where $y = A \cdot [x,u]^T + b$, ϕ non-linearity

Deterministic output: $f(x,u) = g(u)$

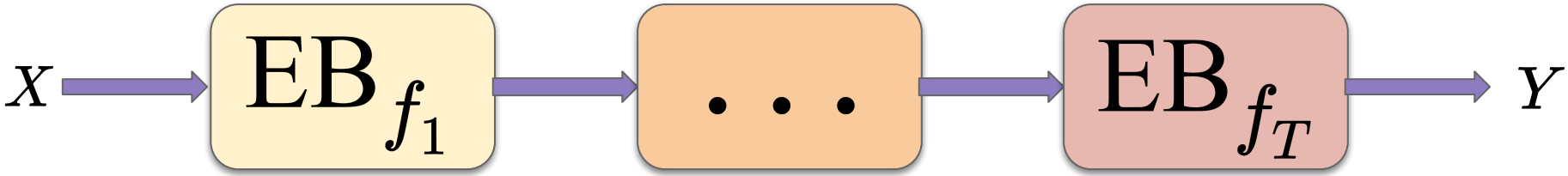
$\rightarrow Y = \mathbb{E}_{X'}[g(X')]$ **deterministic**

Classical warping: $f(x,u) = h(x)$

$\rightarrow Y = h(X)$ **random**

Proposed Architectures

$$Y = \mathrm{EB}_{f_T} \circ \ldots \circ \mathrm{EB}_{f_1}(X)$$



Tasks		
Discriminative	**Generative**	**Predictive**
Y deterministic	Y random	Y random
X random	X noise + code	X random

Approximation Property

➢ *Theoretically, three blocks are enough*

Theorem (Universal Approximation). Let F a continuous map for the convergence in law, mapping measures supported on compact sets. Then ***three EBs*** are necessary to approximate F ***arbitrarily close***: $\forall \epsilon > 0$, there exists three continuous maps ***f, g, h***, such that, for all random vectors X,

$$\text{Wasserstein}_1\left(F(X), \text{EB}_h \circ \Lambda \circ \text{EB}_g \circ \text{EB}_f(X)\right) \leq \epsilon$$

where $\Lambda : X \mapsto (X, U)$ concatenates a uniform random vector.

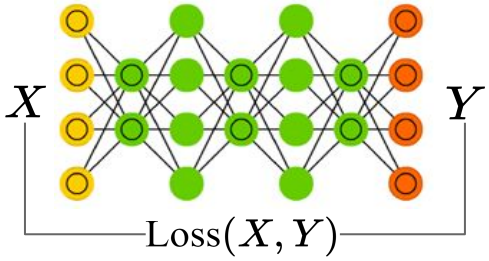
Applications

In practice: f **fully connected**

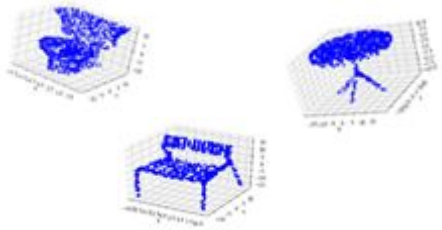
$$f(x, u) = (\phi(y_i))_i$$

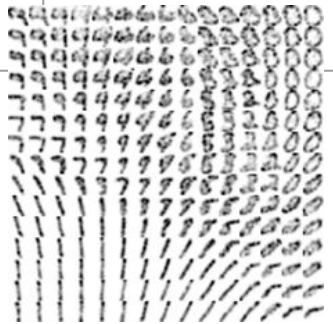
Loss functions:

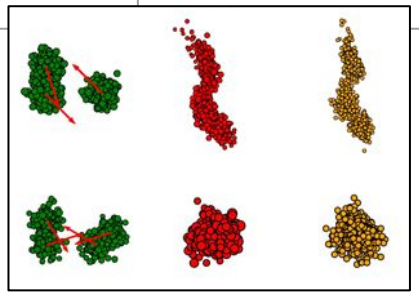
- ***Cross-entropy*** (classification)
- ***Regularized Wasserstein*** (Cuturi, 2013)
- ***Sinkhorn divergence*** (Genevay et al, 2018)



Classification		Generation	Dynamics
MNIST as ***point clouds***	Modelnet40 as ***point clouds***	MNIST as ***point clouds***	Flocking model
99,2% accuracy, 2 EBs	83,5% accuracy, 2 EBs	2 EBs	5 EBs







Conclusion / Open Problems

- **New formalism for stochastic deep architectures**
  - ➢ Probability distributions
  - ➢ Deterministic feature vectors
- **Robustness & approximation power**
- **Perspectives**
  - ➢ Understanding block roles
  - ➢ Investigate translation & rotation equivariance

Poster: #30 Pacific Ballroom today - see you there !