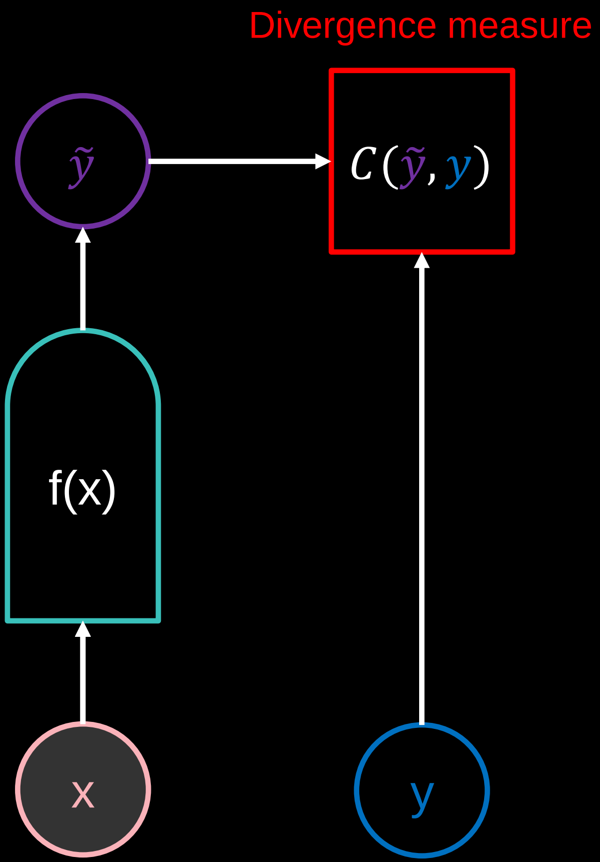
Latent-Variable Energy-Based Models

(LV-EBMs)

Julien Marteen Akay | LinkedIn

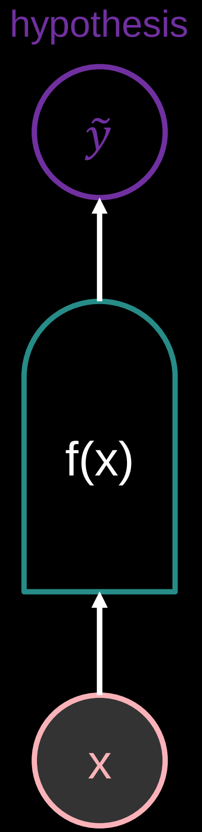
Feed-Forward Nets - Training

- explicit function
- produces one output
- one to many?
  → compromise



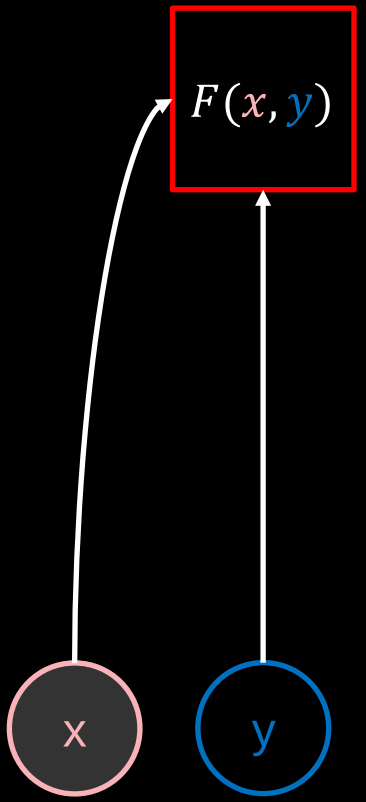
Feed-Forward Nets - Inference

- one finite object as output per observation
- what if…
  - multiple outputs for single input
  - need for constraint satisfaction
  - …



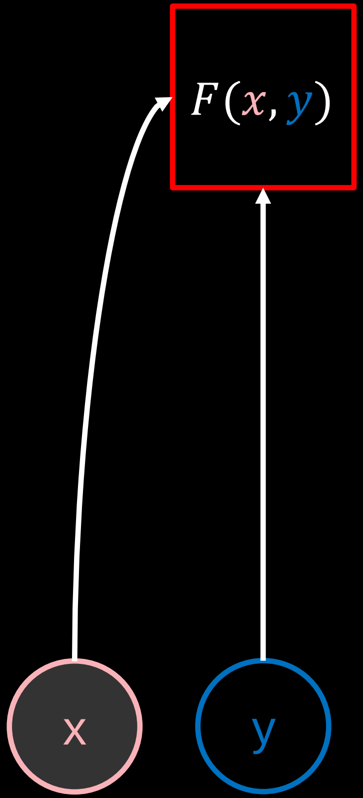
Energy-Based Model – Implicit function

- does not compute y given x
- captures dependencies between x and $\boldsymbol{y}$
- scalar-valued implicit function
- energy function
  - low energy = compatible
  - high energy = incompatible



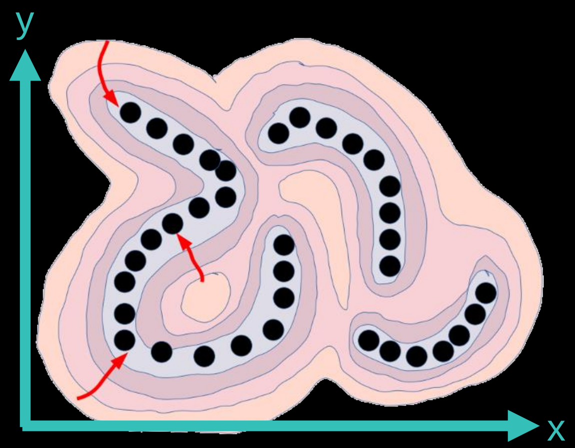
Energy-Based Model – Inference

- given an EBM & observation x
- find $\check{\boldsymbol{y}} = \underset{\boldsymbol{y}}{\operatorname{argmin}}\, F(x, \boldsymbol{y})$
- inference involves optimization process



Energy-Based Model – Inference

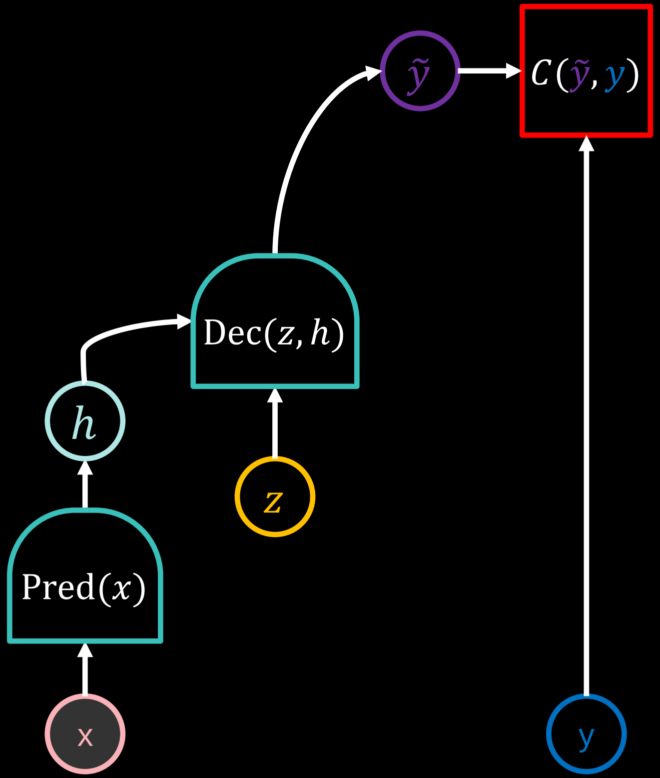
- given EBM & observation x
- find $\check{y} = \underset{y}{\operatorname{argmin}}\, F(x, y)$
- inference involves optimization process
- multiple low energy y for x
- F with respect to y continuous and smooth?
  → gradient-based method for inference



[1]

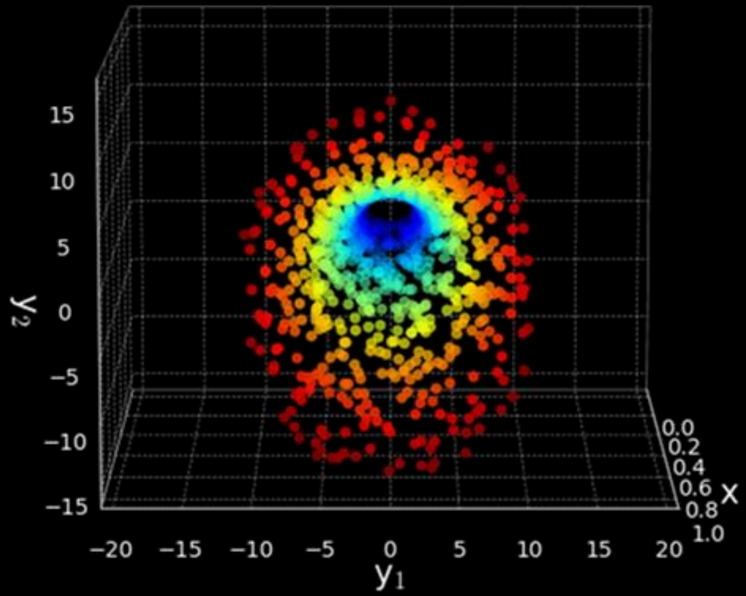
Latent-Variable Energy-Based Model

- latent variable z parameterizes set of predictions Y
- model infers how to use z
- z (ideally) represents independent explanatory factors of variation of the prediction



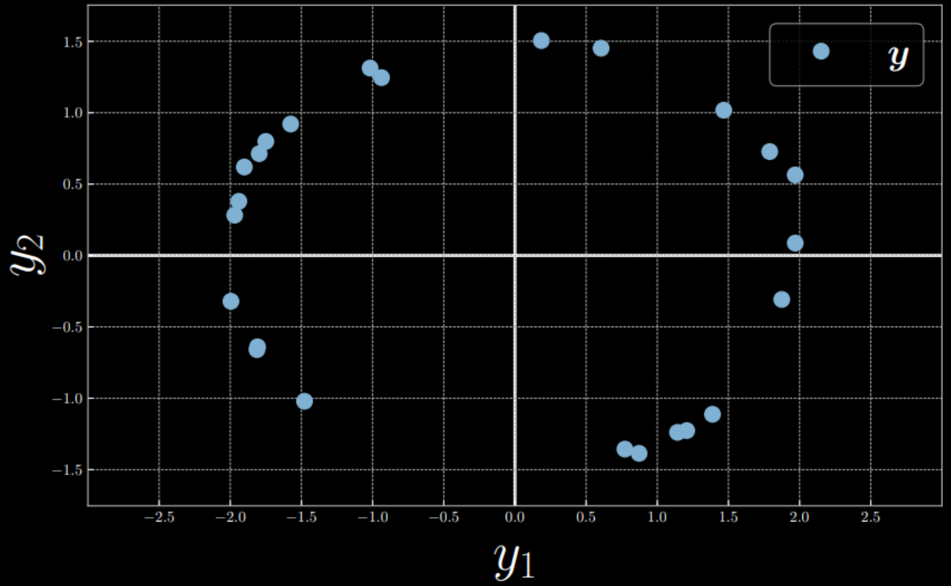
Toy data set

- $\boldsymbol{y} = \begin{bmatrix} \rho_1(x)\cos(\theta) + \varepsilon \\ \rho_2(x)\sin(\theta) + \varepsilon \end{bmatrix}$
- $\rho(x) = \begin{bmatrix} \alpha x + \beta(1-x) \\ \beta x + \alpha(1-x) \end{bmatrix} \cdot e^{2x}$
- $(\alpha, \beta) = (1.5, 2)$
- $x \sim \mathcal{U}(0,1)$
- $\theta \sim \mathcal{U}(0, 2\pi)$
- $\varepsilon \sim \mathcal{N}(0, 20^{-2})$



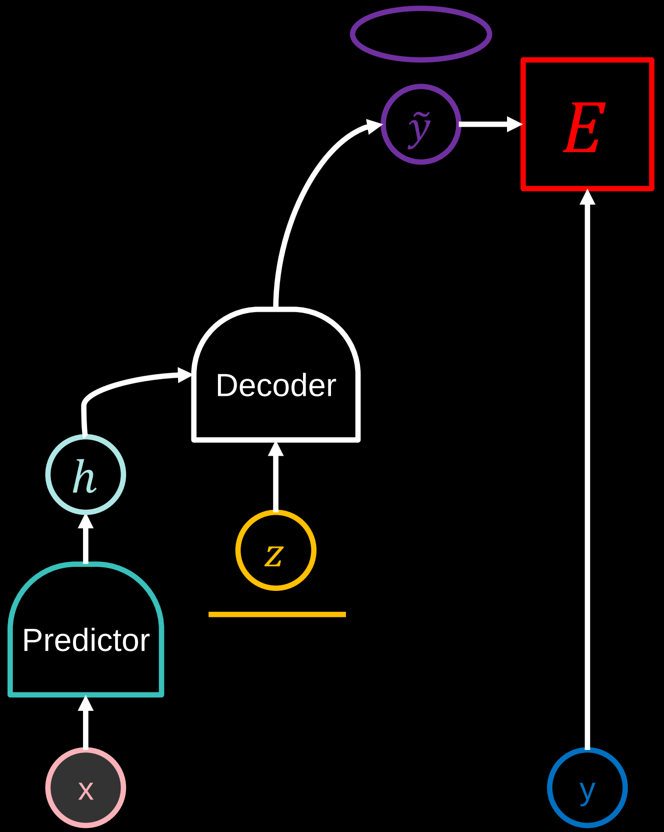
Toy data set

- $\boldsymbol{y} = \begin{bmatrix} \rho_1(x)\cos(\theta) + \varepsilon \\ \rho_2(x)\sin(\theta) + \varepsilon \end{bmatrix}$
- $\rho(x) = \begin{bmatrix} \alpha x + \beta(1-x) \\ \beta x + \alpha(1-x) \end{bmatrix} \cdot e^{2x}$
- $(\alpha, \beta) = (1.5, 2)$
- $x \sim \cancel{\mathcal{U}(0,1)}\ 0$
- $\theta \sim \mathcal{U}(0, 2\pi)$
- $\varepsilon \sim \mathcal{N}(0, 20^{-2})$



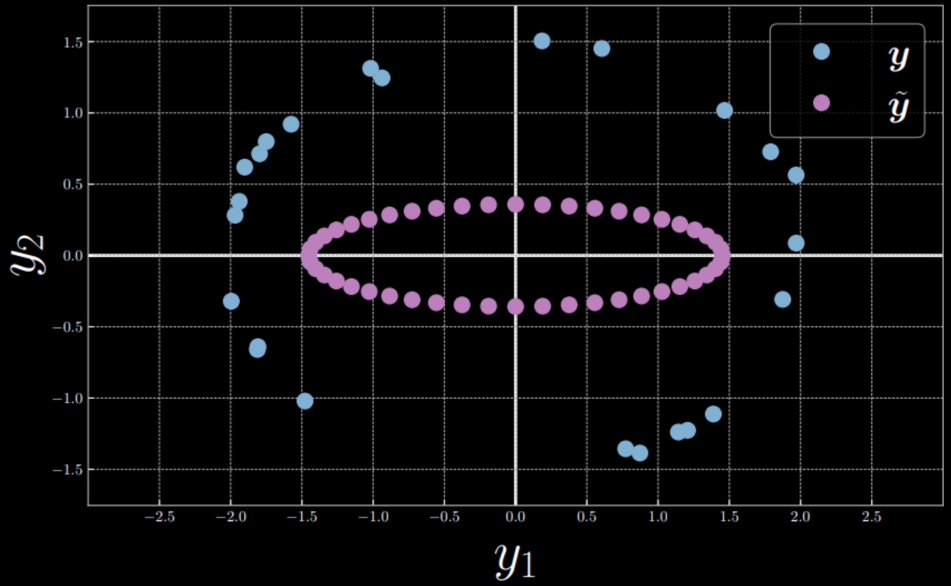
LV-EBM - Architecture

- $\tilde{y} = f_\phi(x) \odot g(z)$
- $f_\phi, g: \mathbb{R} \to \mathbb{R}^2$
- $g(z) = \begin{bmatrix} \cos(z) \\ \sin(z) \end{bmatrix}$
- $z = \left[0 : \frac{\pi}{24} : 2\pi\right[$



LV-EBM – Untrained

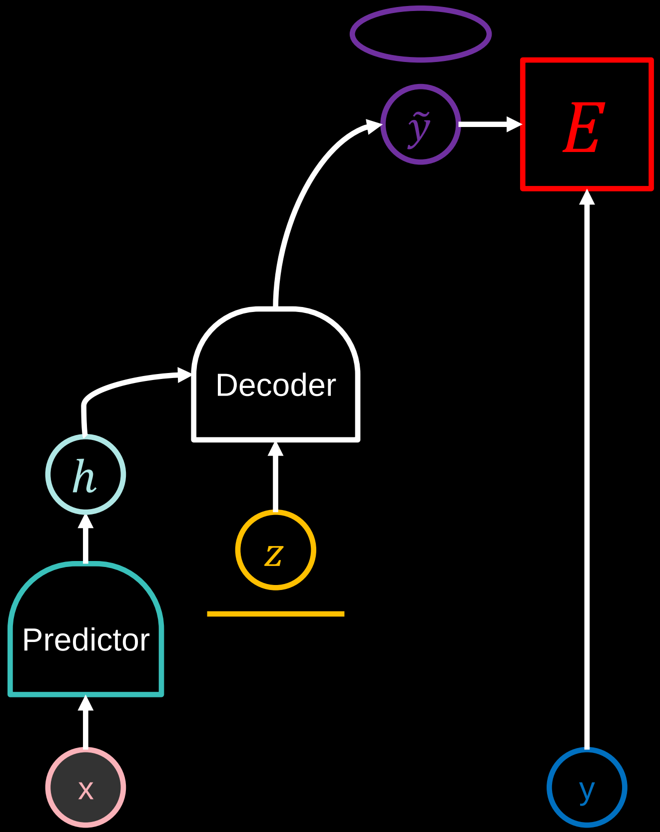
- $\tilde{y} = f_\phi(x) \odot g(z)$
- $f_\phi, g \colon \mathbb{R} \to \mathbb{R}^2$
- $g(z) = \begin{bmatrix} \cos(z) \\ \sin(z) \end{bmatrix}$
- $x = 0$
- $z = \left[0 : \frac{\pi}{24} : 2\pi\right[$



[2]

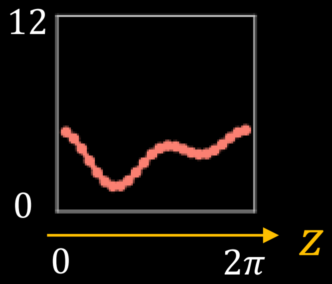
LV-EBM – Energy function

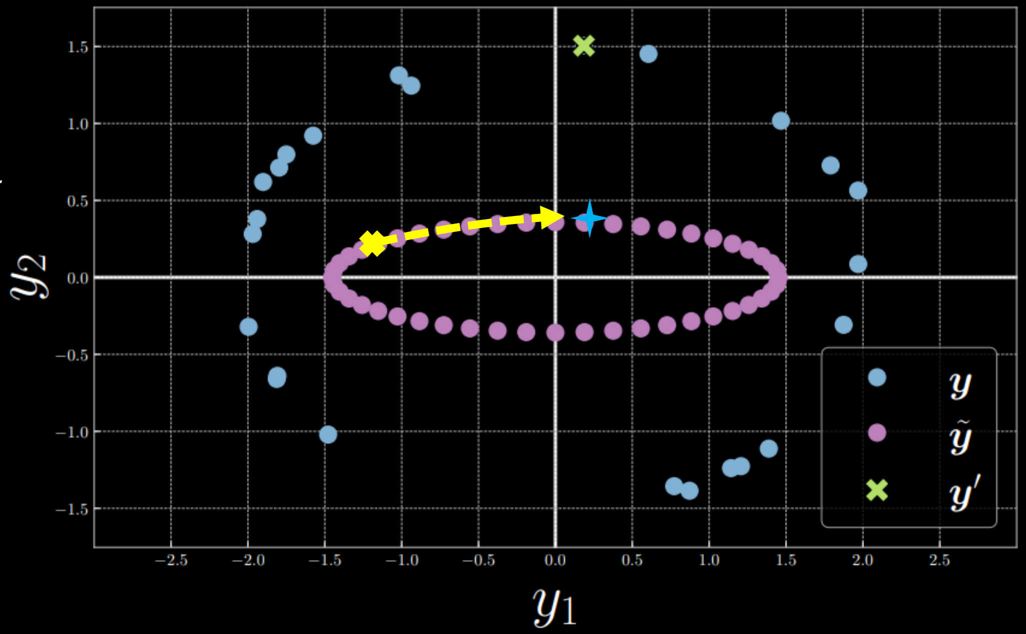
- $\tilde{y} = f_\phi(x) \odot g(z)$
- $f_\phi, g : \mathbb{R} \rightarrow \mathbb{R}^2$
- $g(z) = \begin{bmatrix} \cos(z) \\ \sin(z) \end{bmatrix}$
- $z = \left[0 : \frac{\pi}{24} : 2\pi\right[$
- $E(x, y, z) = \|y - \tilde{y}\|_2^2$
- $= \left\|y - f_\phi(x) \odot g(z)\right\|_2^2$



LV-EBM – Free Energy

- eliminate z, e.g. by minimization
- $F_\infty(x, \boldsymbol{y}) = \min_z E(x, \boldsymbol{y}, z)$
- $\check{\boldsymbol{y}} = \text{argmin}_y F(x, \boldsymbol{y})$
- *„Zero-Temperature-Limit of the free energy“*

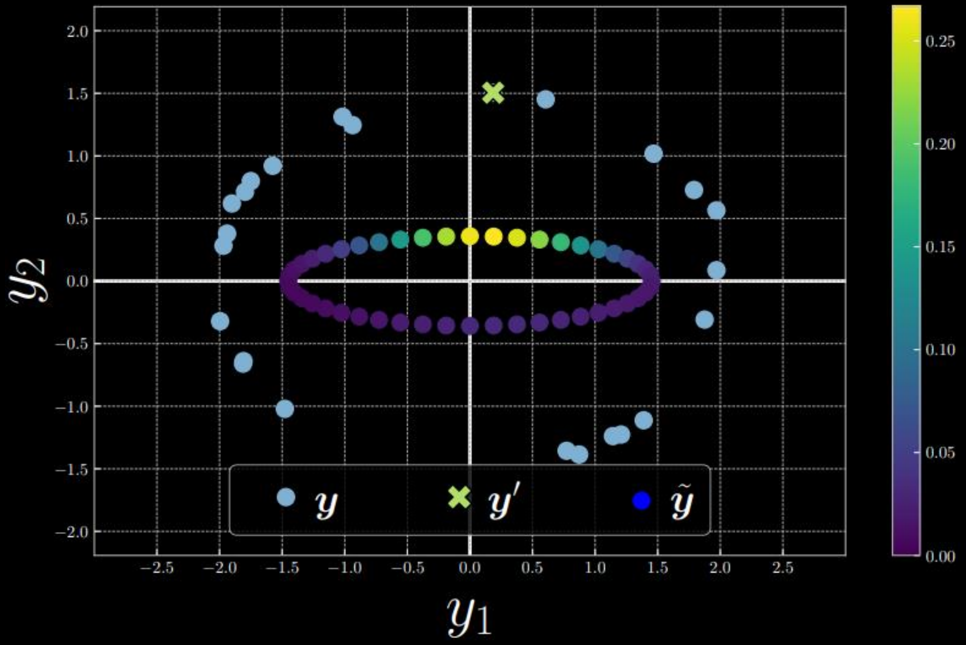




[2]

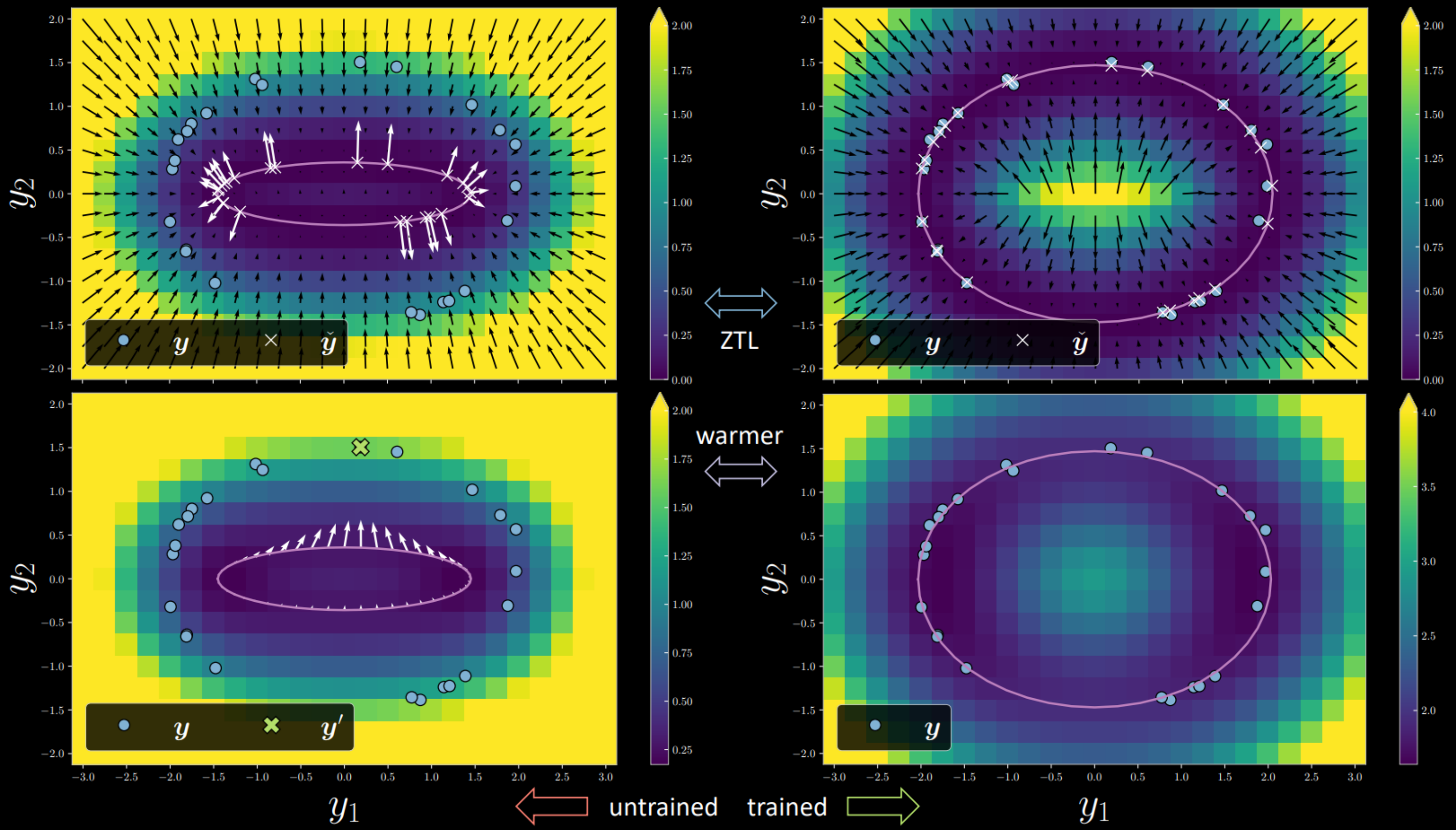
LV-EBM – Free Energy

- $F_\beta(x, y) = -\frac{1}{\beta} \log \frac{1}{|\mathcal{Z}|} \int_{\mathcal{Z}} e^{-\beta E(x,y,z)} \mathrm{d}z$
- *„softmin“*



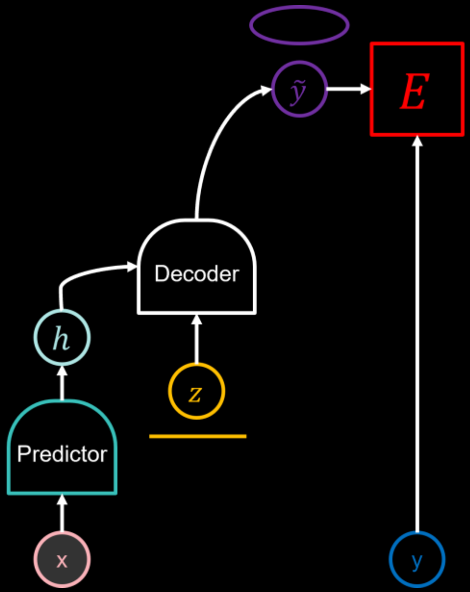
[2]

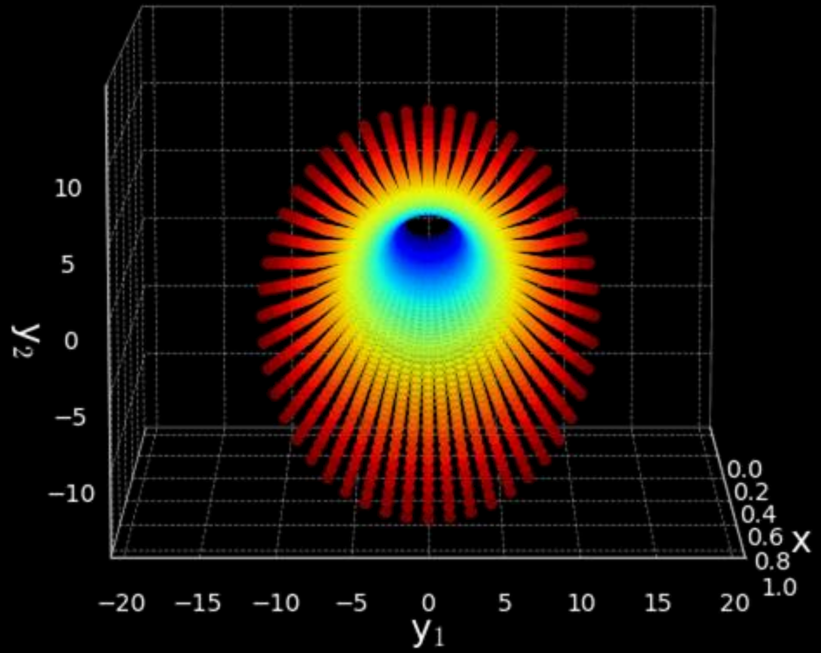
Min (=ZTL) vs. Softmin



[2]

Evaluation – Model manifold





- $x = \left[0 : \frac{1}{50} : 1\right]$ $\quad z = \left[0 : \frac{\pi}{24} : 2\pi\right[$

That’s it. Thanks!

ANY QUESTIONS?

Sources

[1] A Path Towards Autonomous Machine Intelligence | OpenReview - https://openreview.net/forum?id=BZ5a1r-kVsf

[2] New York University (NYU) – Deep Learning (https://atcold.github.io/NYU-DLSP21/) - https://github.com/Atcold/pytorch-Deep-Learning/blob/master/slides/12%20-%20EBM.pdf

[3] Model-Predictive Policy Learning with Uncertainty Regularization for Driving in Dense Traffic | Project website https://sites.google.com/view/model-predictive-driving/

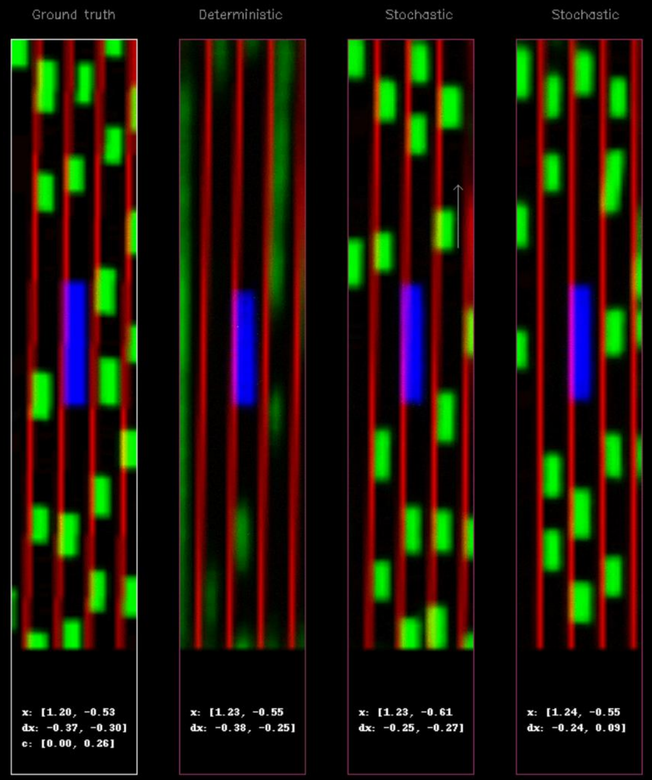
[4] The Impact of chatGPT talks (2023) - Keynote address by Prof. Yann LeCun (NYU/Meta) | YouTube https://www.youtube.com/watch?v=vyqXLJsmsrk&t=2704s

Appendix

- Prediction: Determinstic vs. Non-Deterministic
- The Future of Large Language Models

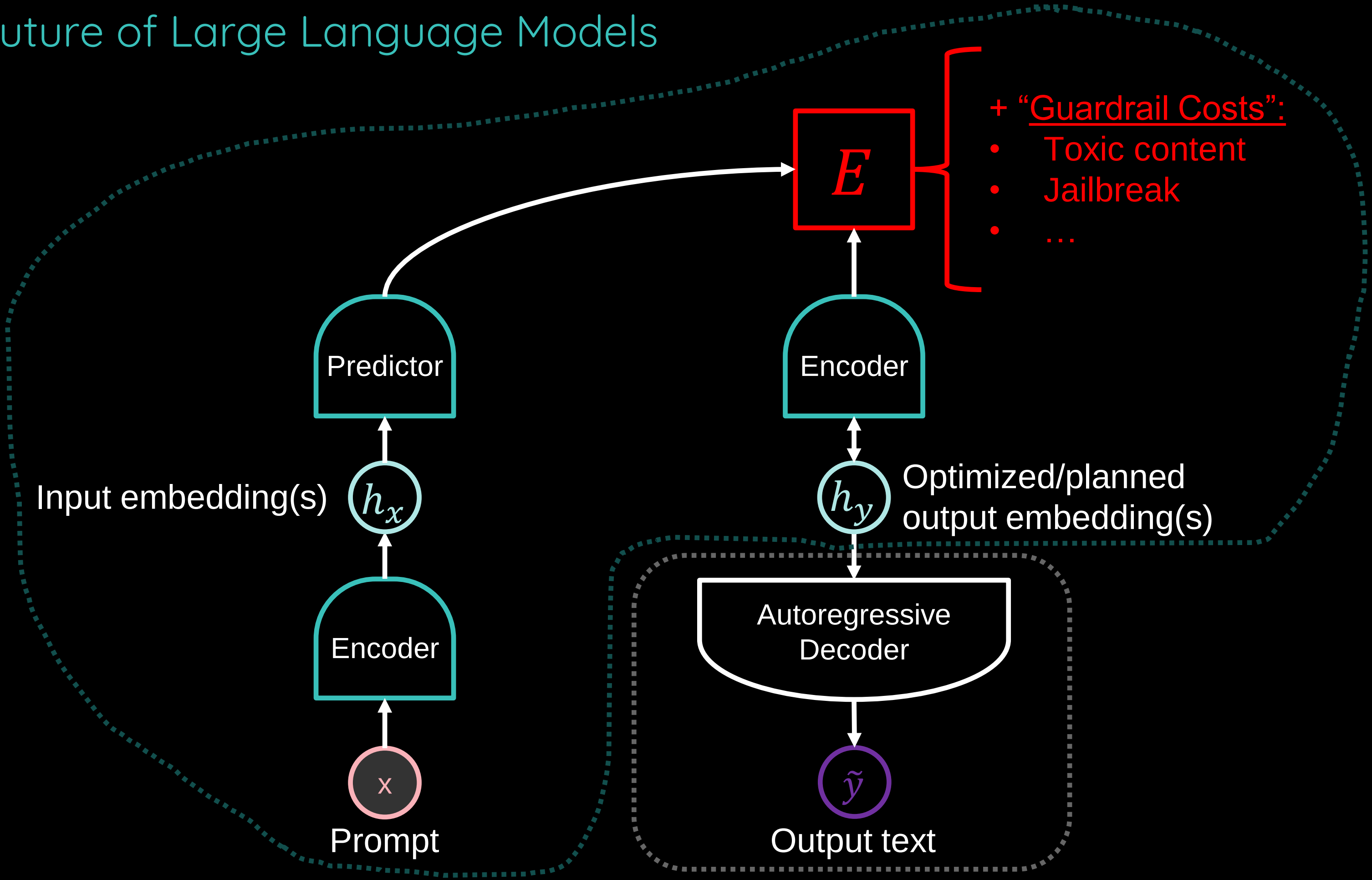
Prediction: Deterministic vs. Non-Deterministic

- deterministic model produces blurry predictions



[3]

The Future of Large Language Models



[4]