Universität Rostock Traditio et Innovatio

Neural Density Estimation

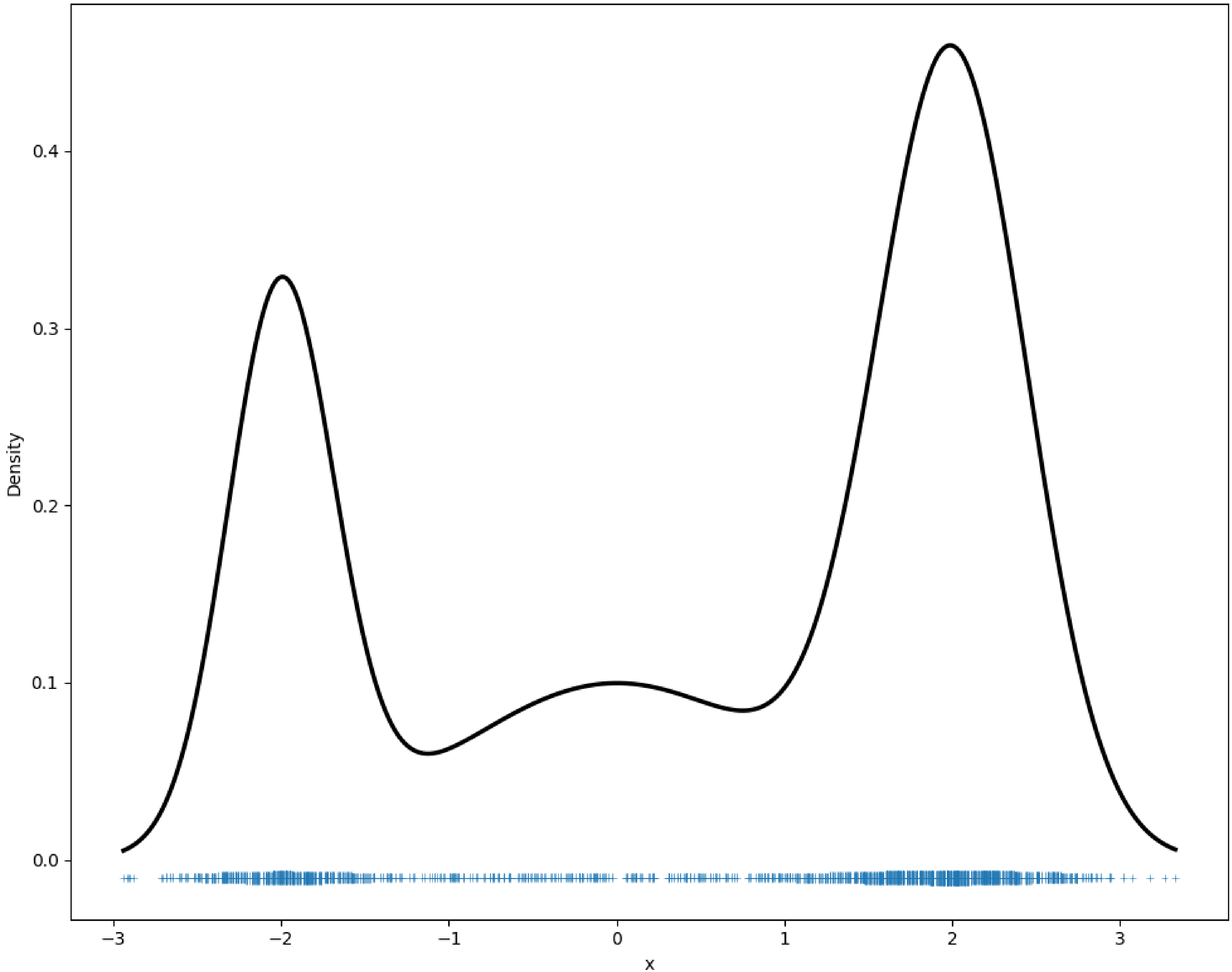
Introduction into Normalising Flows and Masked Autoregressive Flow

Daniel Decker
University of Rostock

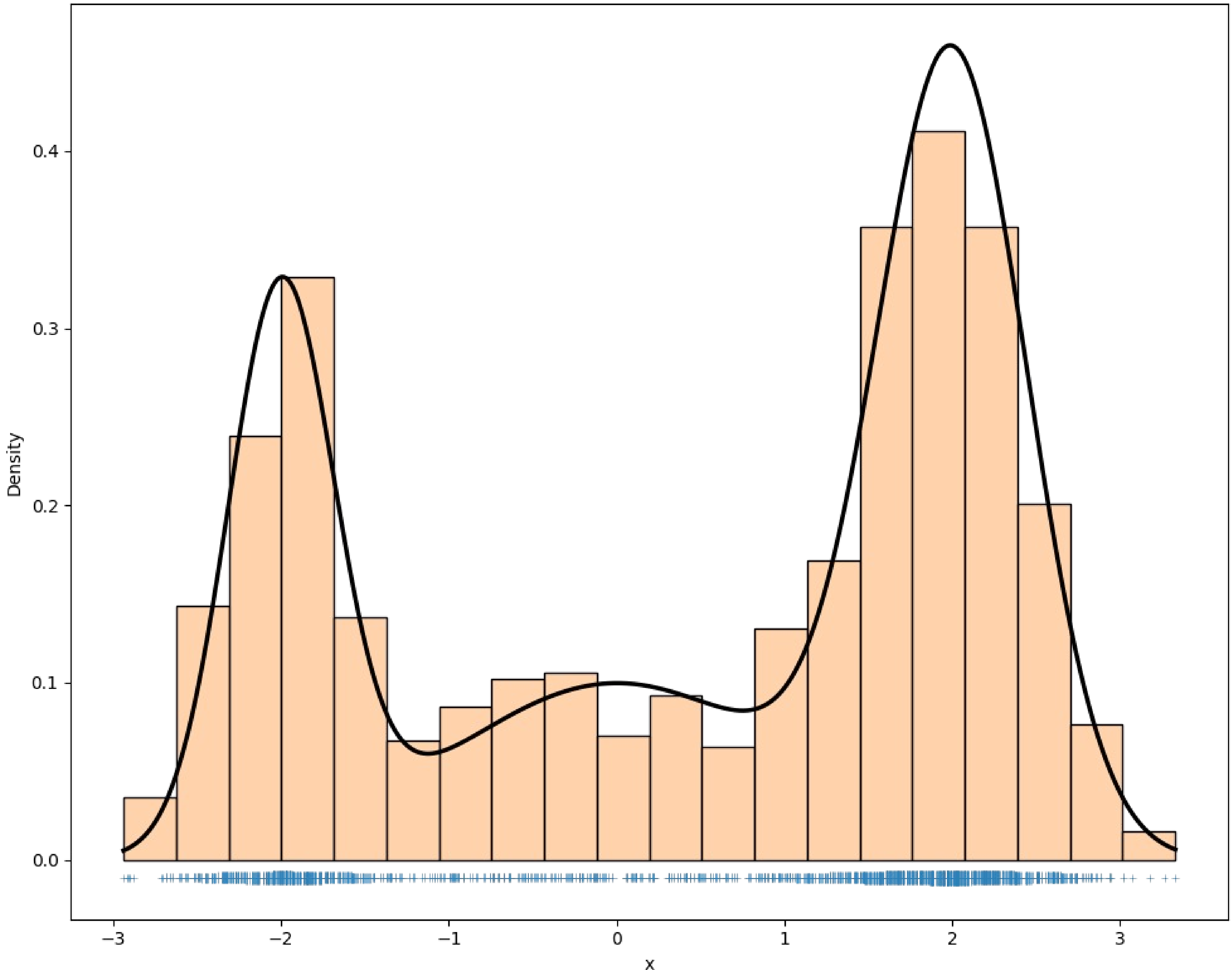
Content

- Density Estimation & Masked Autoregressive Flows
- Evaluation of Masked Autoregressive Flow:
  - Density Estimation with artificial distributions
  - Sampling quality with a real world dataset

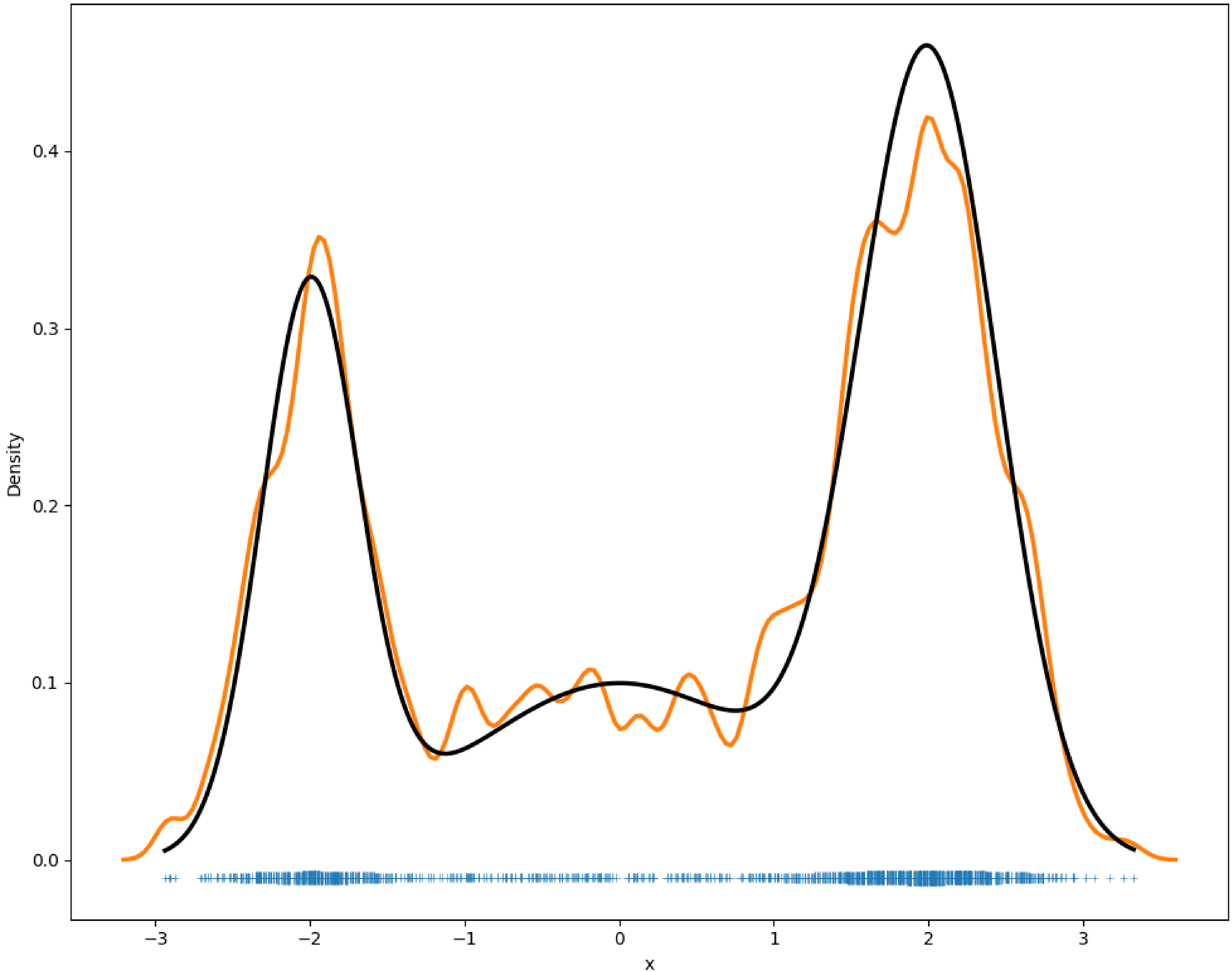
Intro: Density Estimation



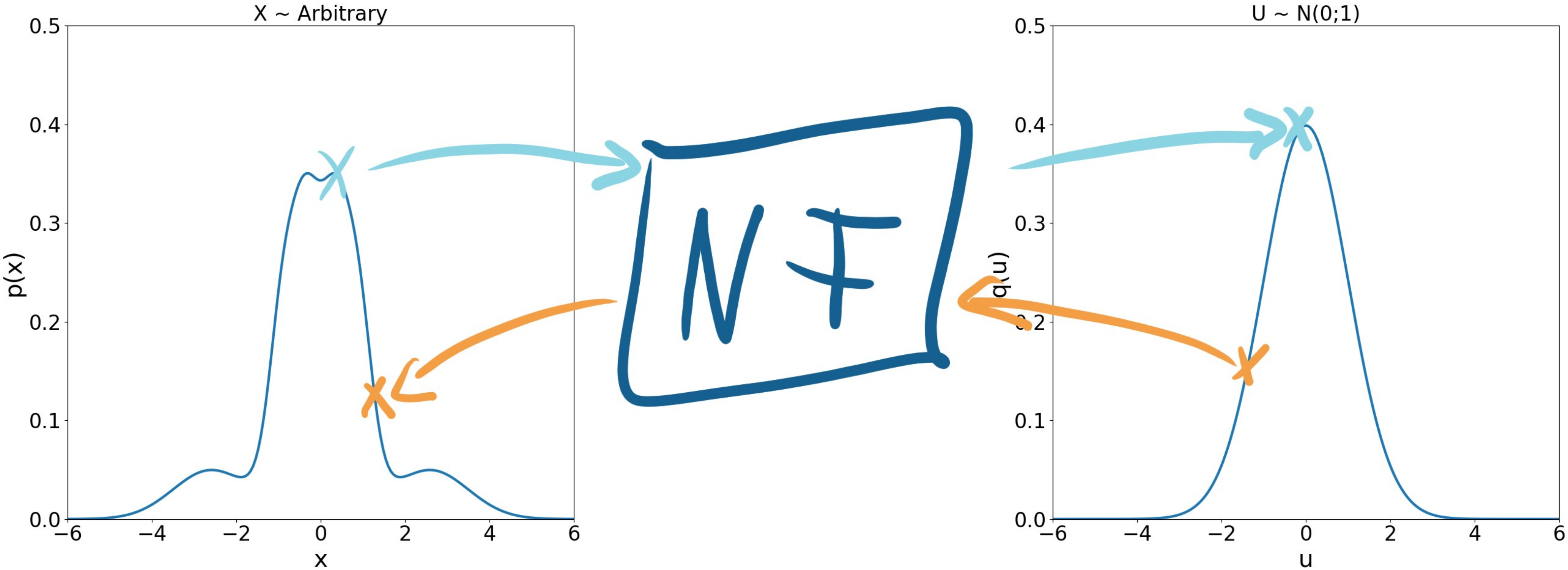
Intro: Density Estimation - Histogram



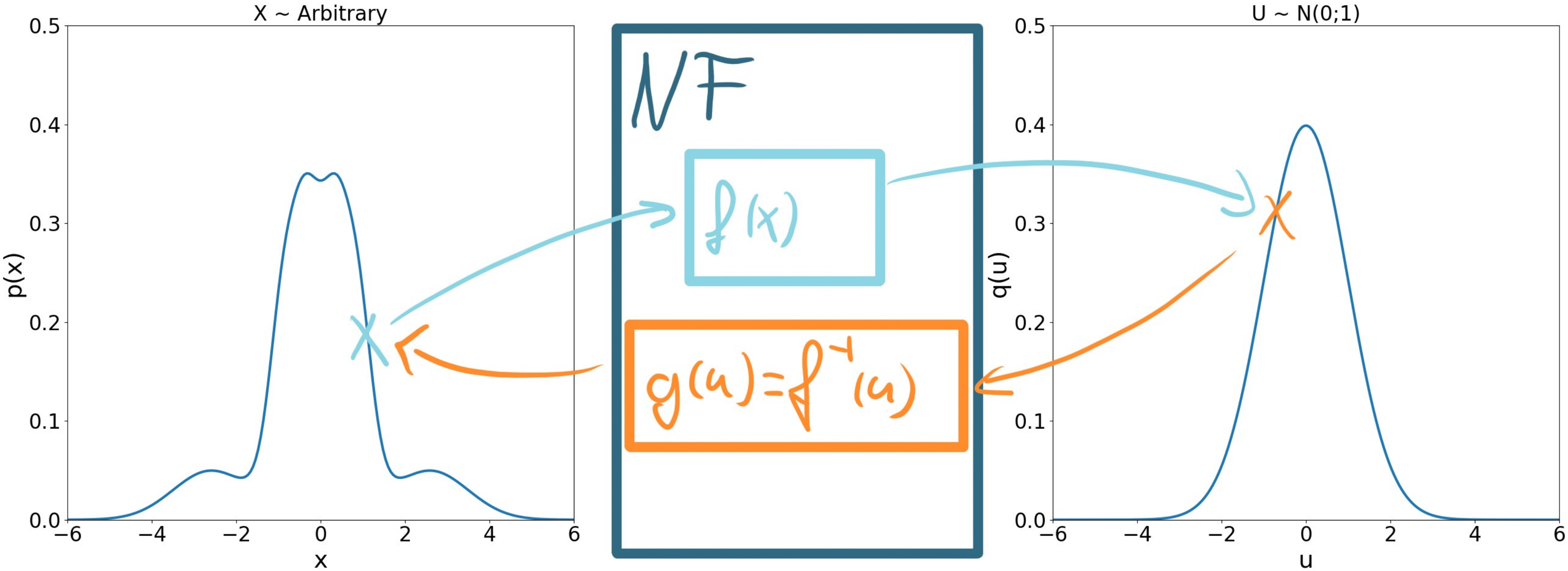
Intro: Density Estimation - KDE



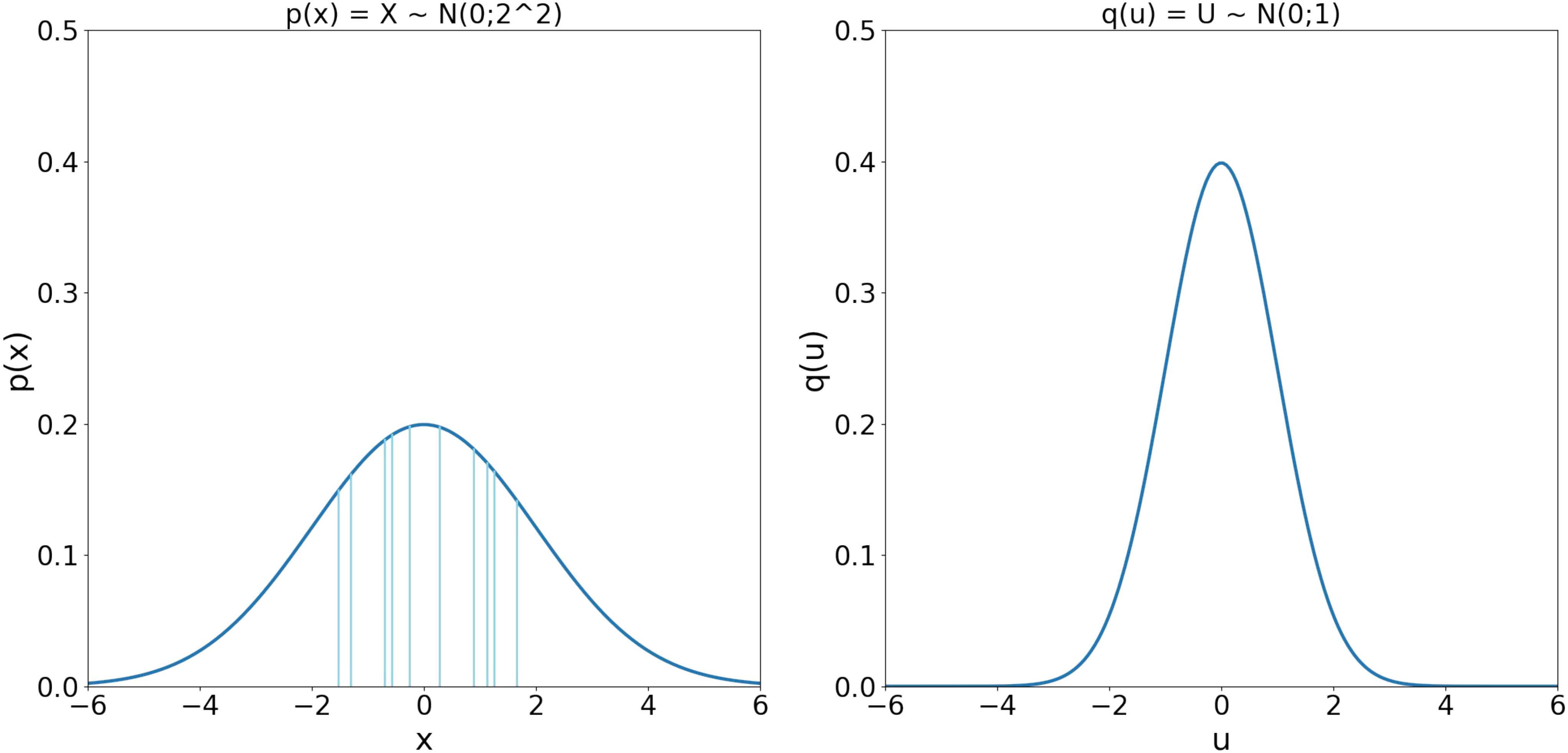
Intro: Normalising Flows



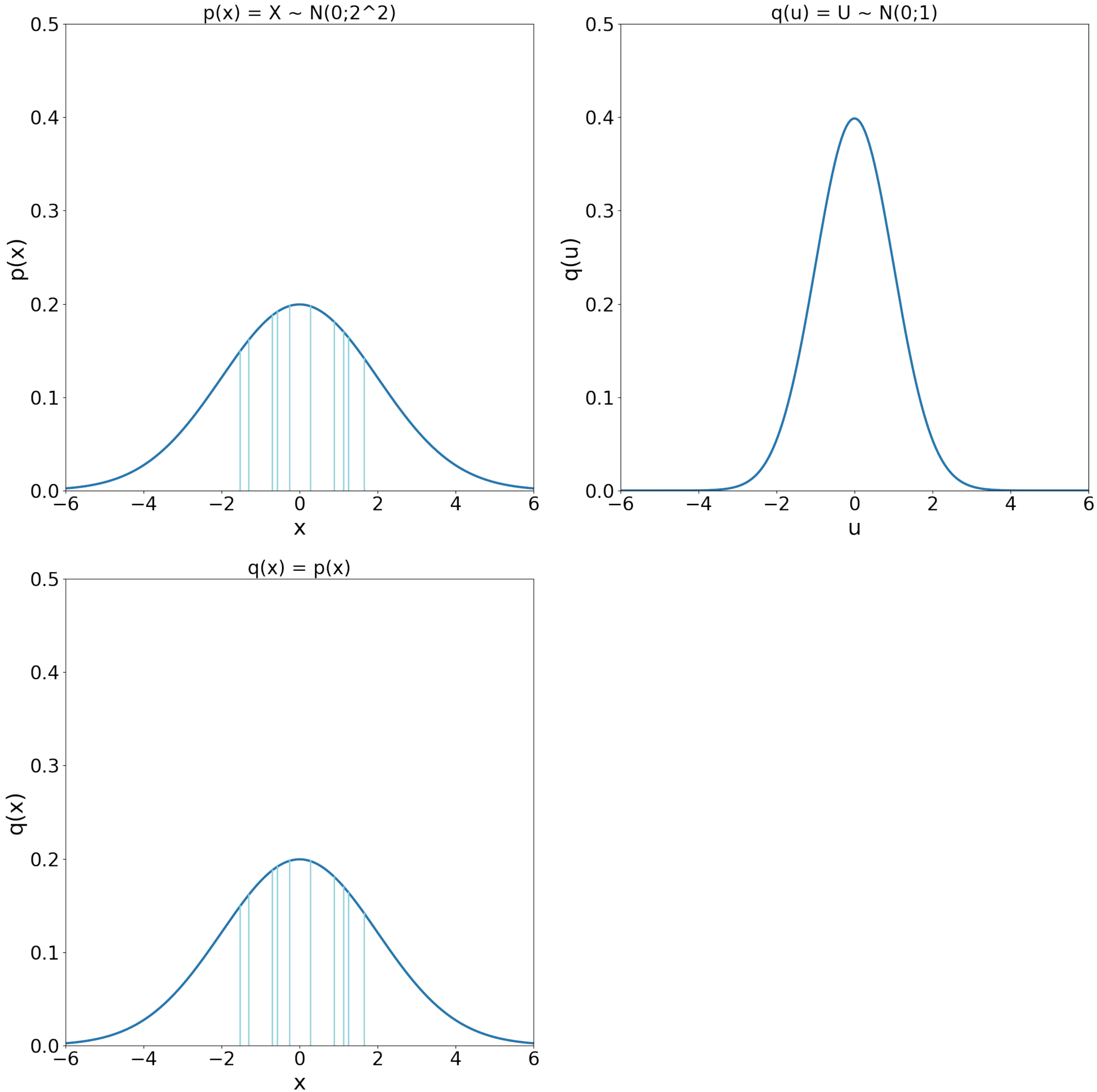
Intro: Normalising Flows



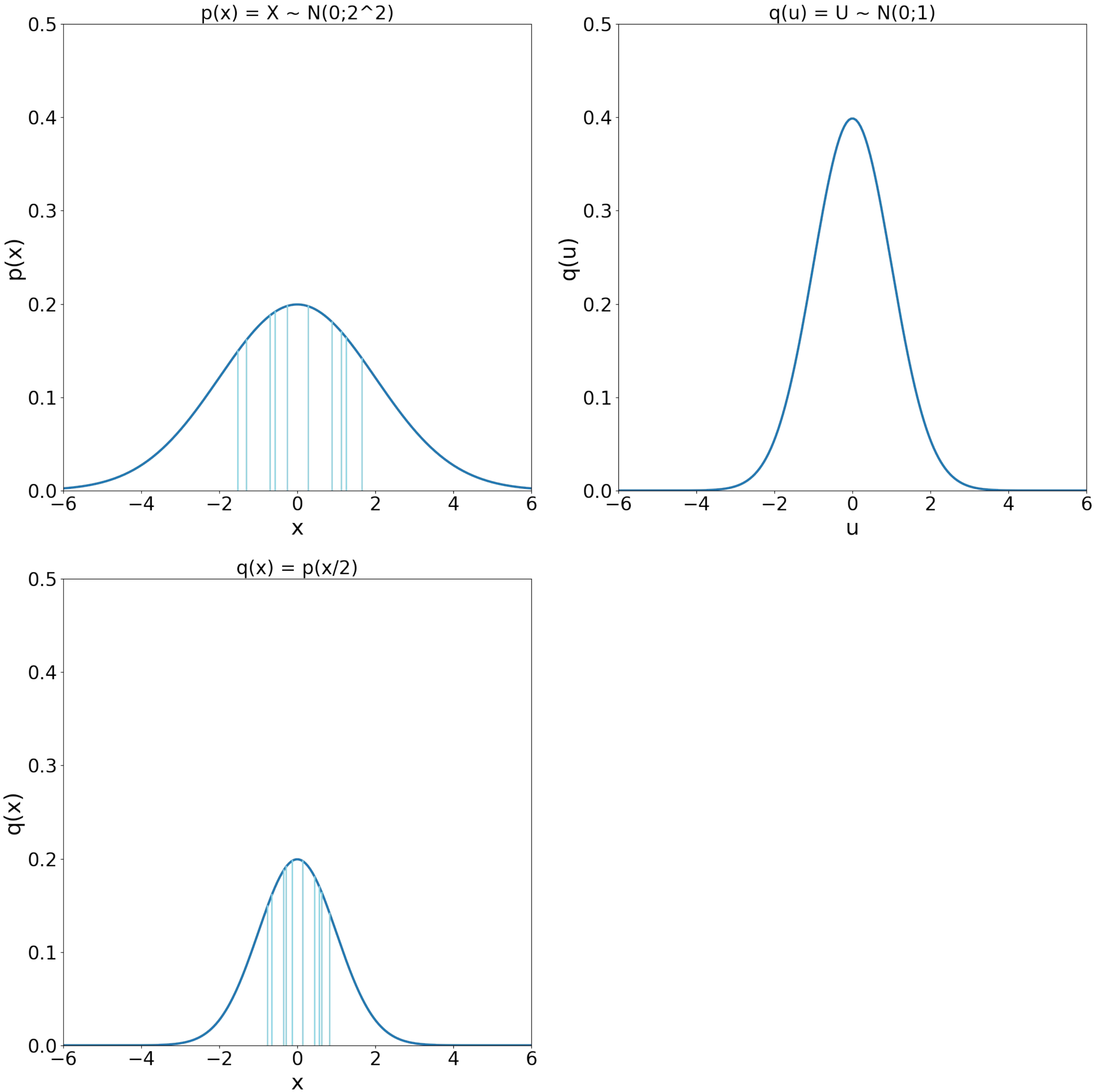
Intro: 1D Transformation



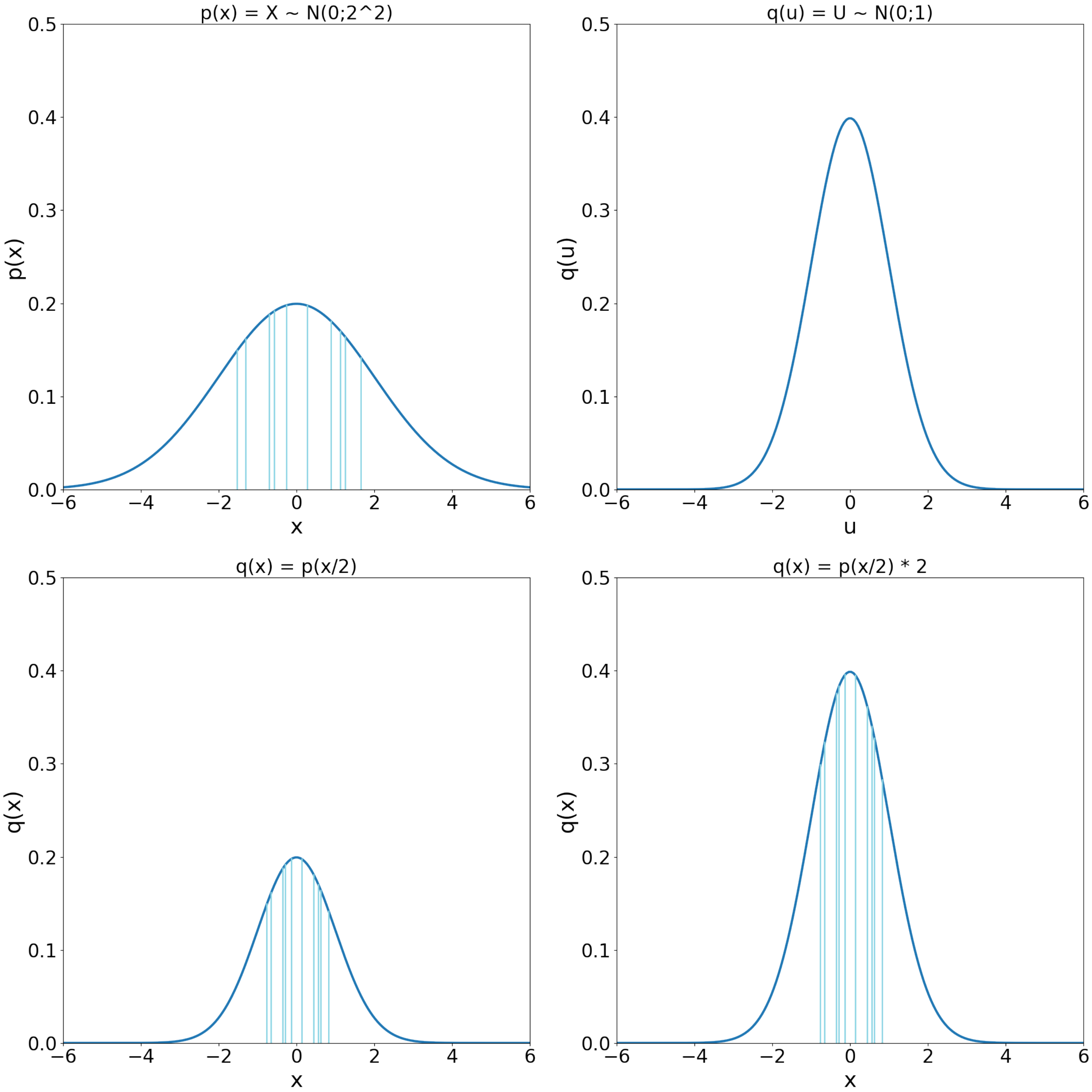
Intro: 1D Transformation



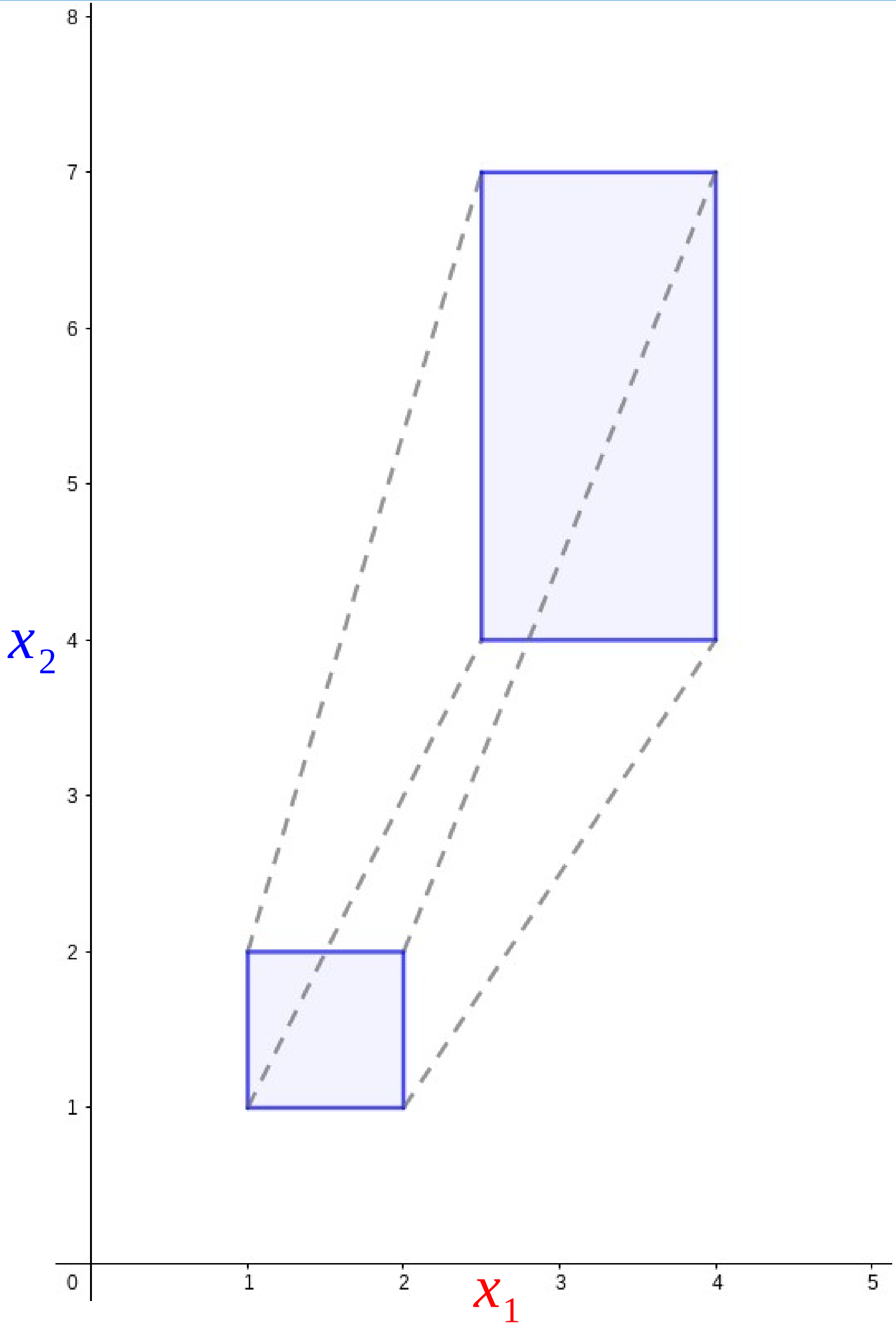
Intro: 1D Transformation



Intro: 1D Transformation



Intro: Jacobian Matrix and Determinant



$$f(x) = \begin{matrix} f_1(x_1) = \frac{3}{2}x_1 + 1 \\ f_2(x_2) = 3x_2 + 1 \end{matrix}$$

$$\text{Jacobian}(f(x)) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{bmatrix}$$

$$\text{Jacobian}(f(x)) = \begin{bmatrix} \frac{3}{2} & 0 \\ 0 & 3 \end{bmatrix}$$

$$\det \text{Jacobian}(f(x)) = \frac{9}{2}$$

Intro: Normalising Flows

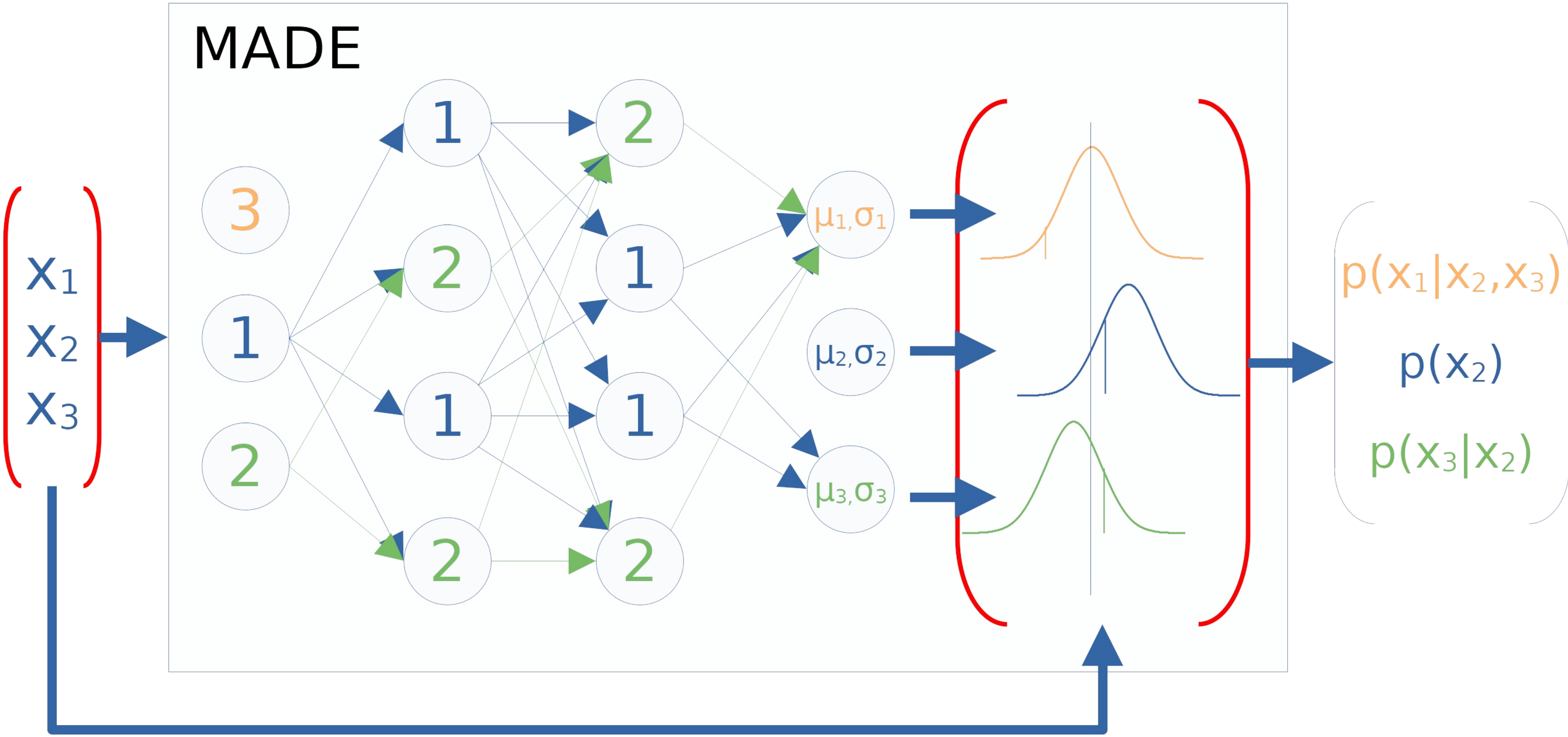
- Normalising Flow formula:

$$p(\boldsymbol{x}) = q(f(\boldsymbol{x}))\left|det\left(\frac{\partial f}{\partial \boldsymbol{x}}\right)\right|$$

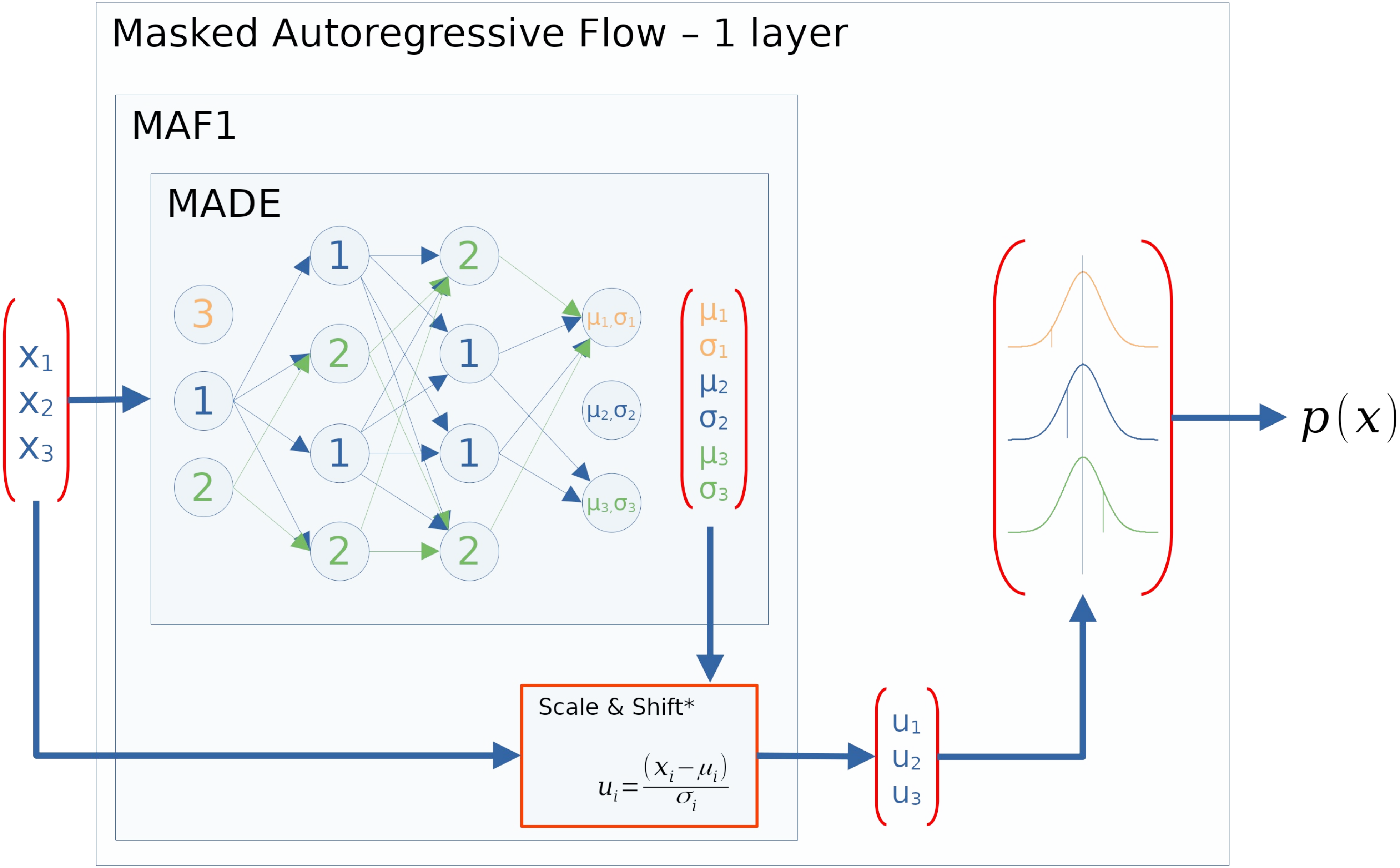
- Optimisation target: maximum likelihood
- Deep models can be build
- Deep models have increased expressiveness

Intro: Autoregression

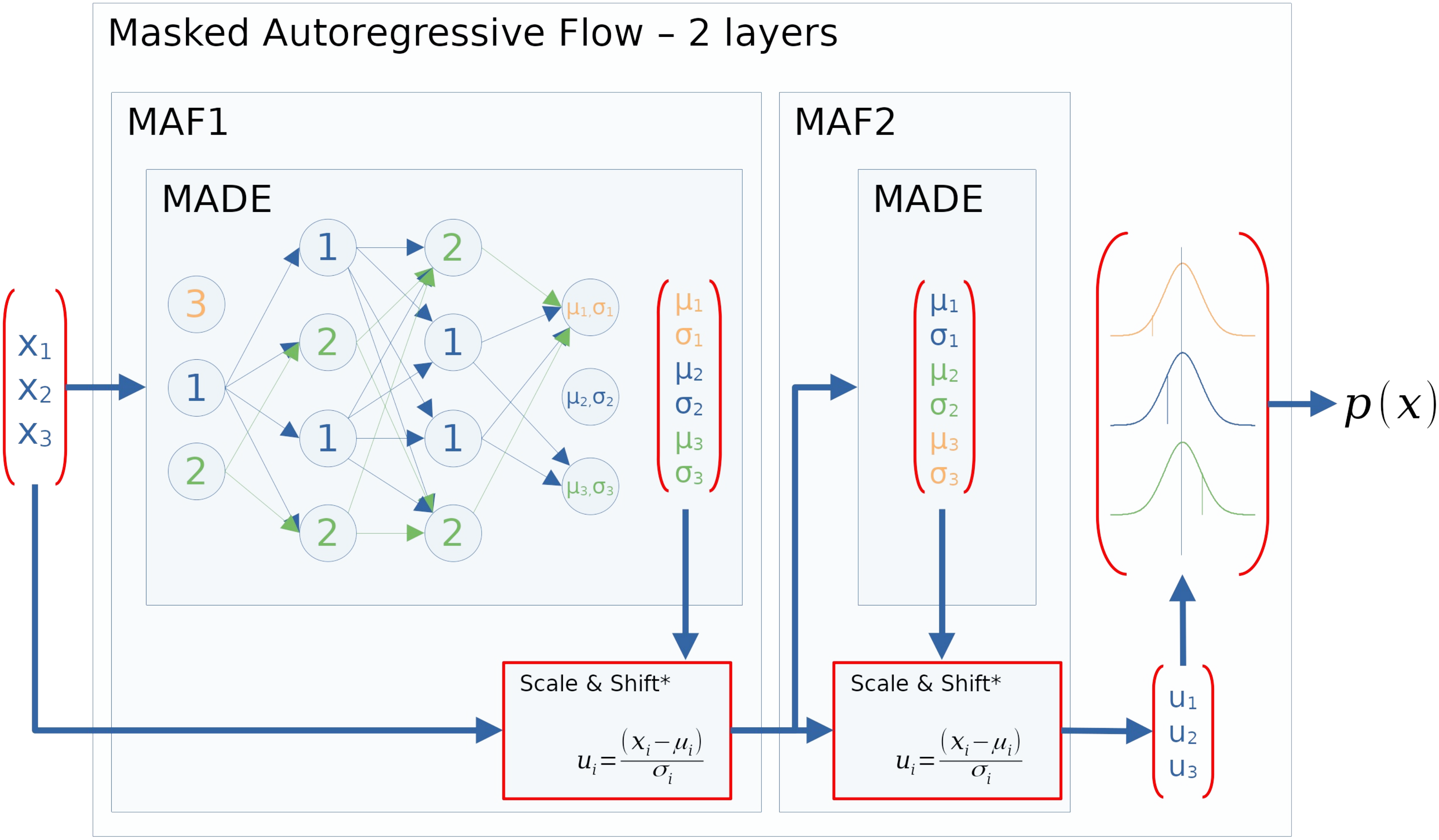
$$p(\boldsymbol{x}) = \prod_{1}^{D} p(x_i | \boldsymbol{x}_{1:i})$$



Intro: Masked Autoregressive Flow



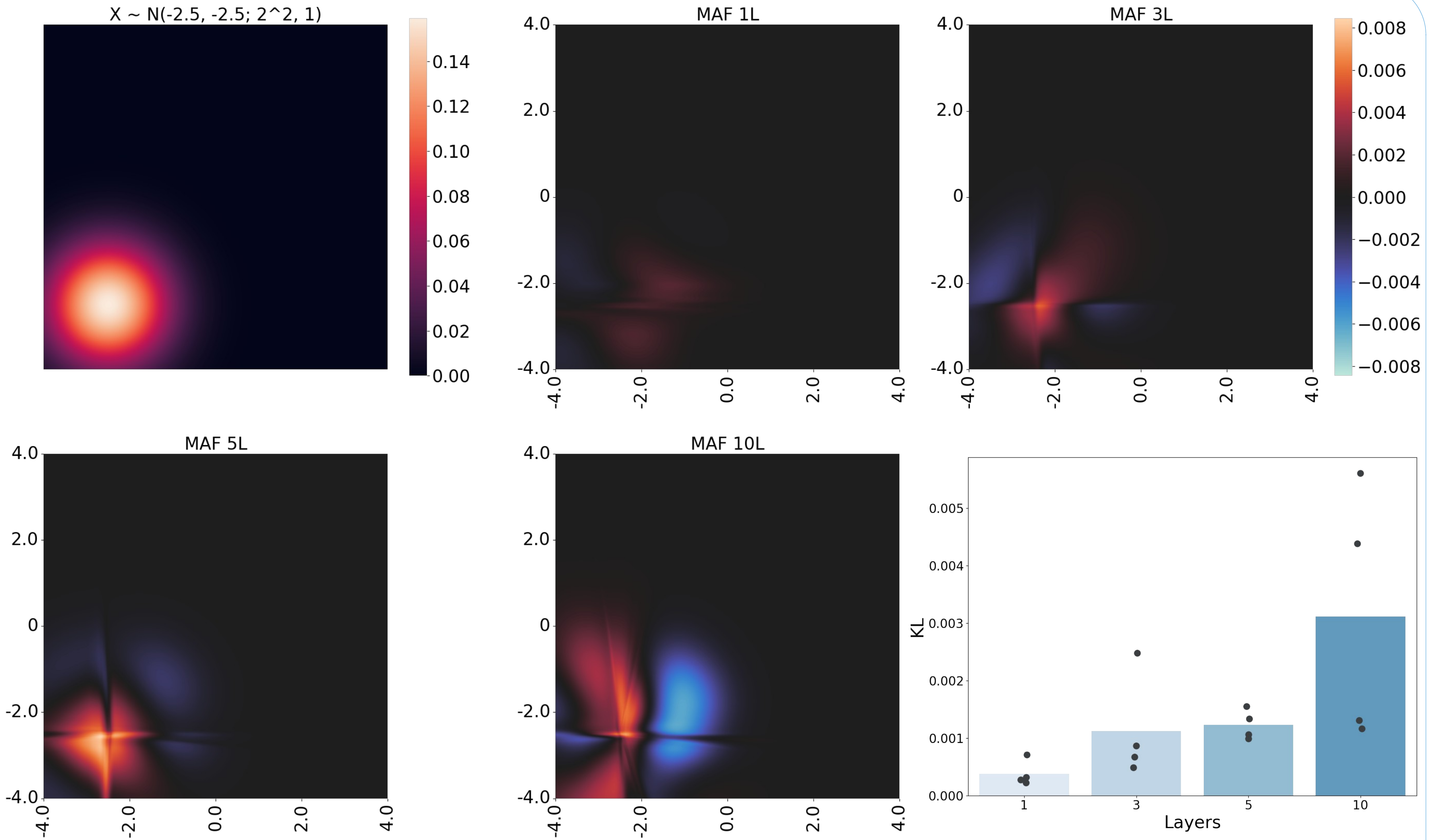
Intro: Deep Masked Autoregressive Flow



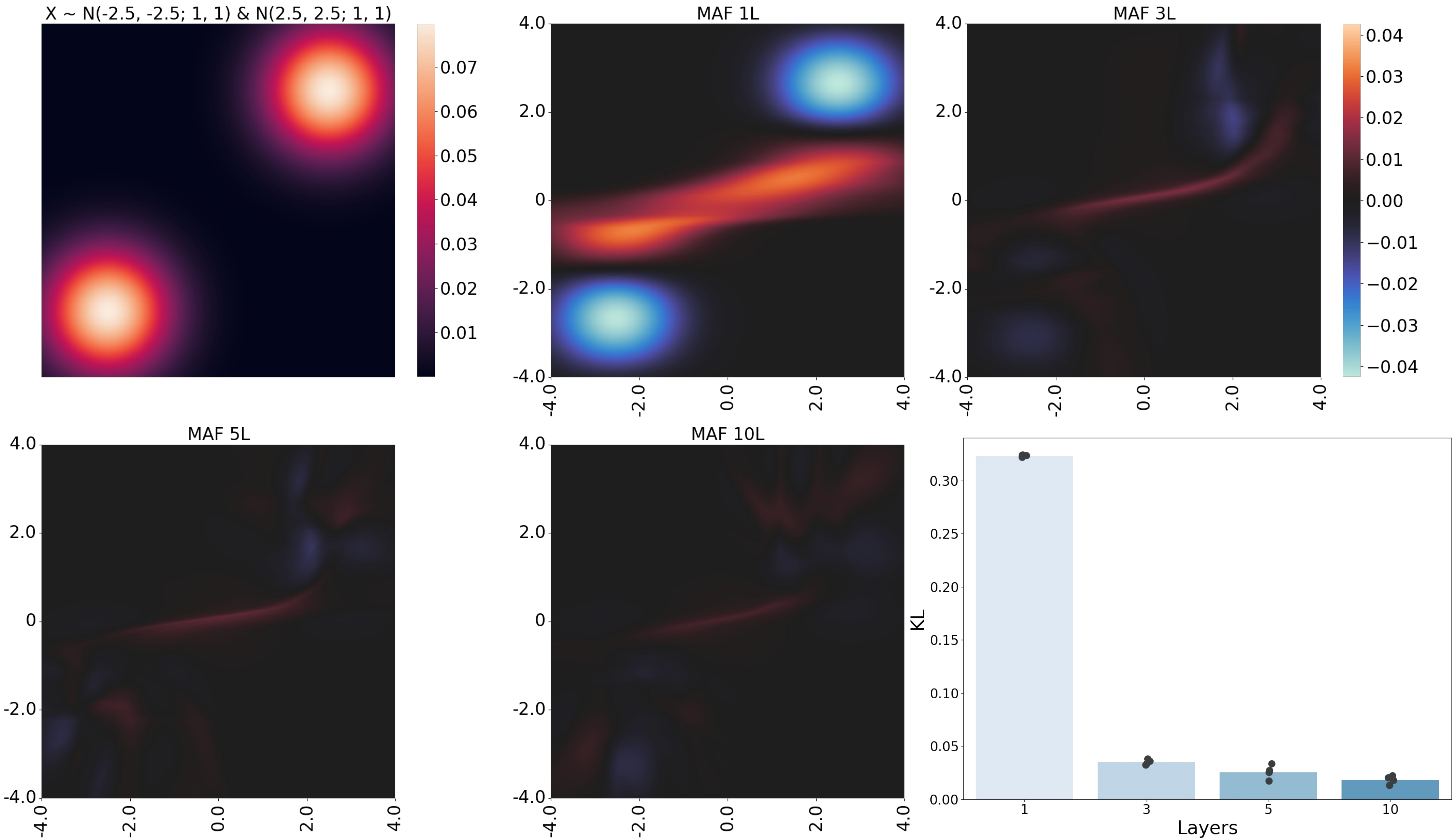
Practical Concerns

- Density Estimation with artificial distributions
  - Measure: KL-divergence
  - Experiments:
    - Lower dimensions (2D)
    - Higher dimensions (10D)
- Sampling quality with a real world dataset
  - Measure: Accuracy of trained classifiers
  - Experiment:
    - High dimensional (50D)

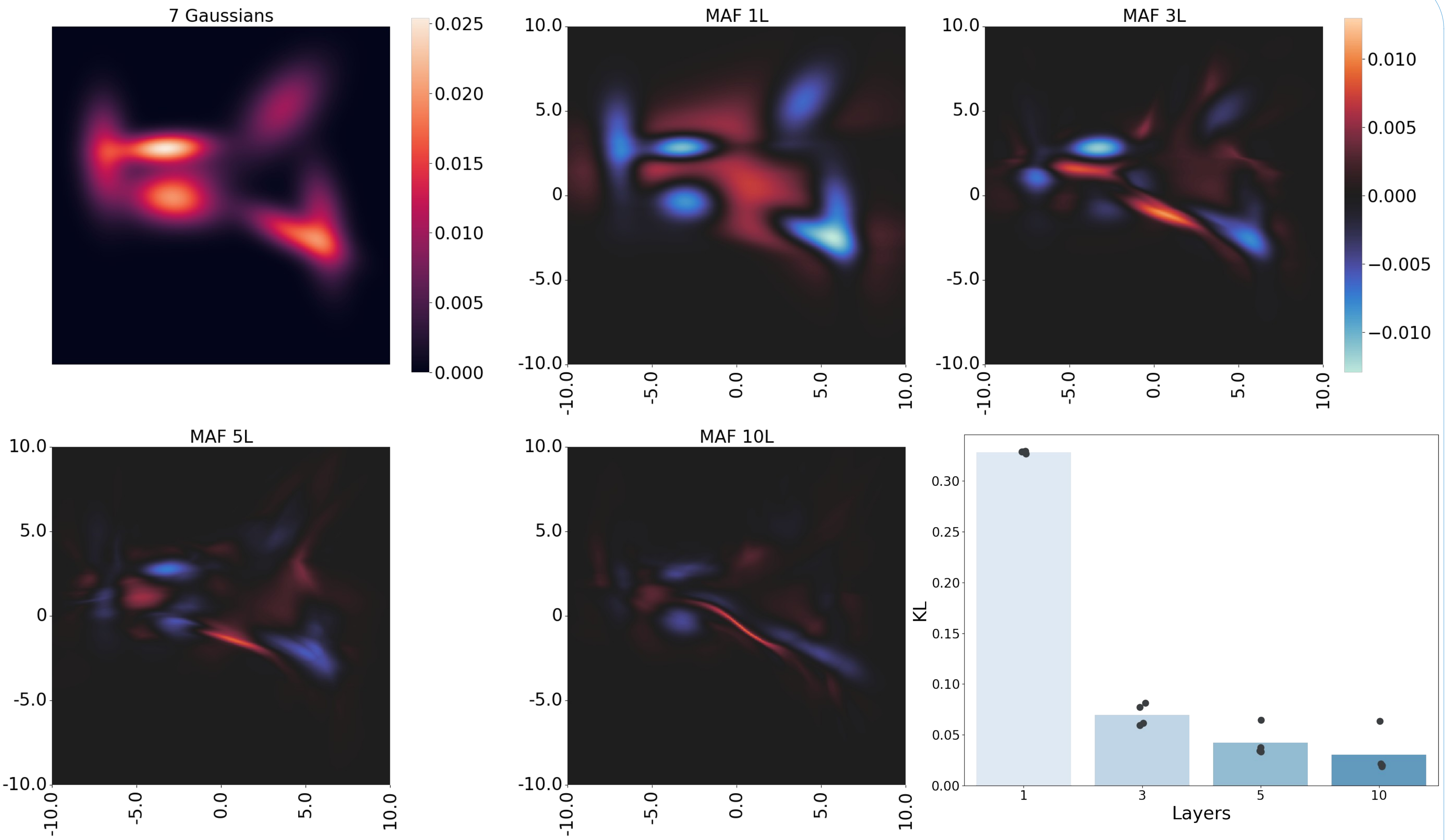
Density Estimation - 2D: 1 Gaussian



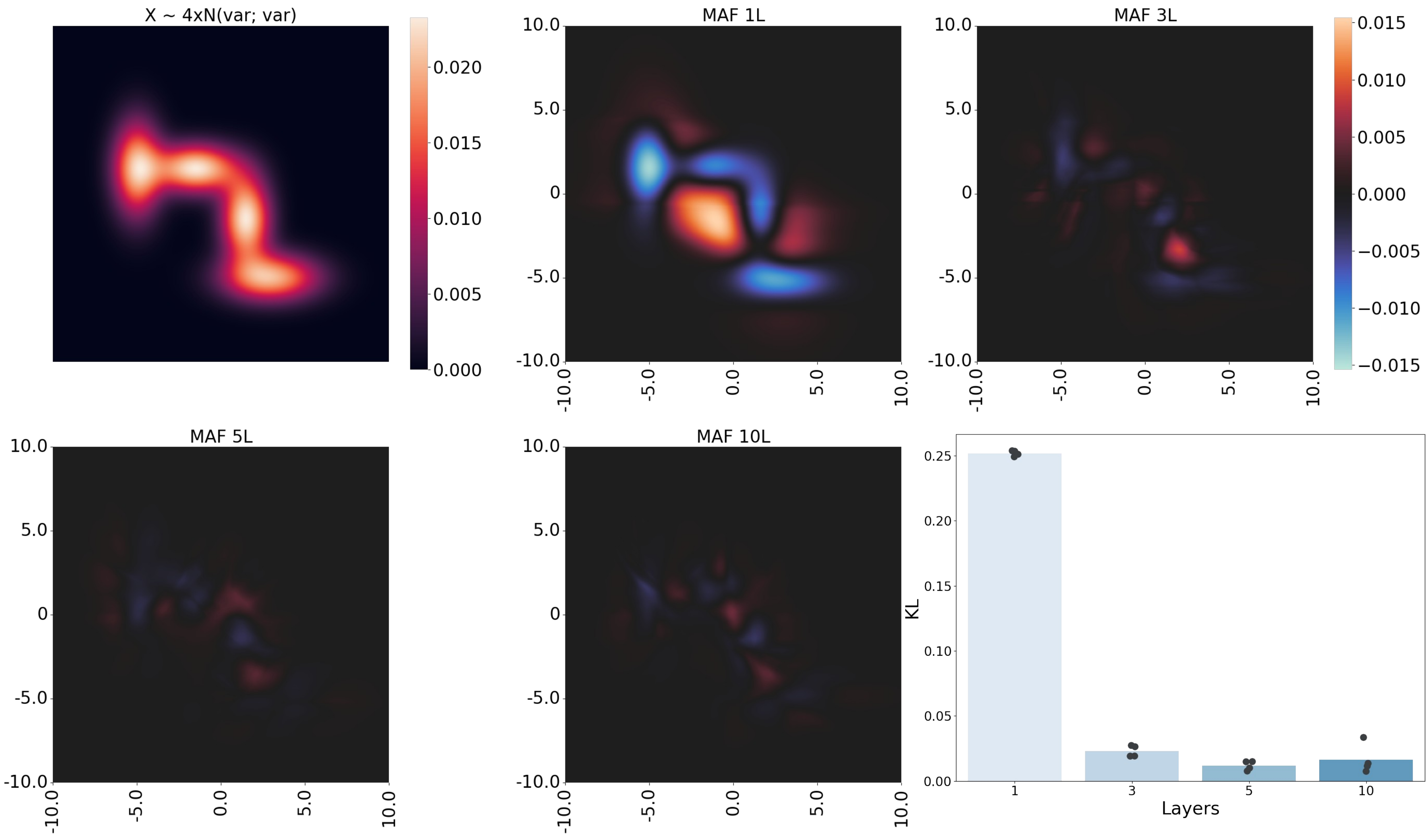
Density Estimation - 2D: 2 Gaussians



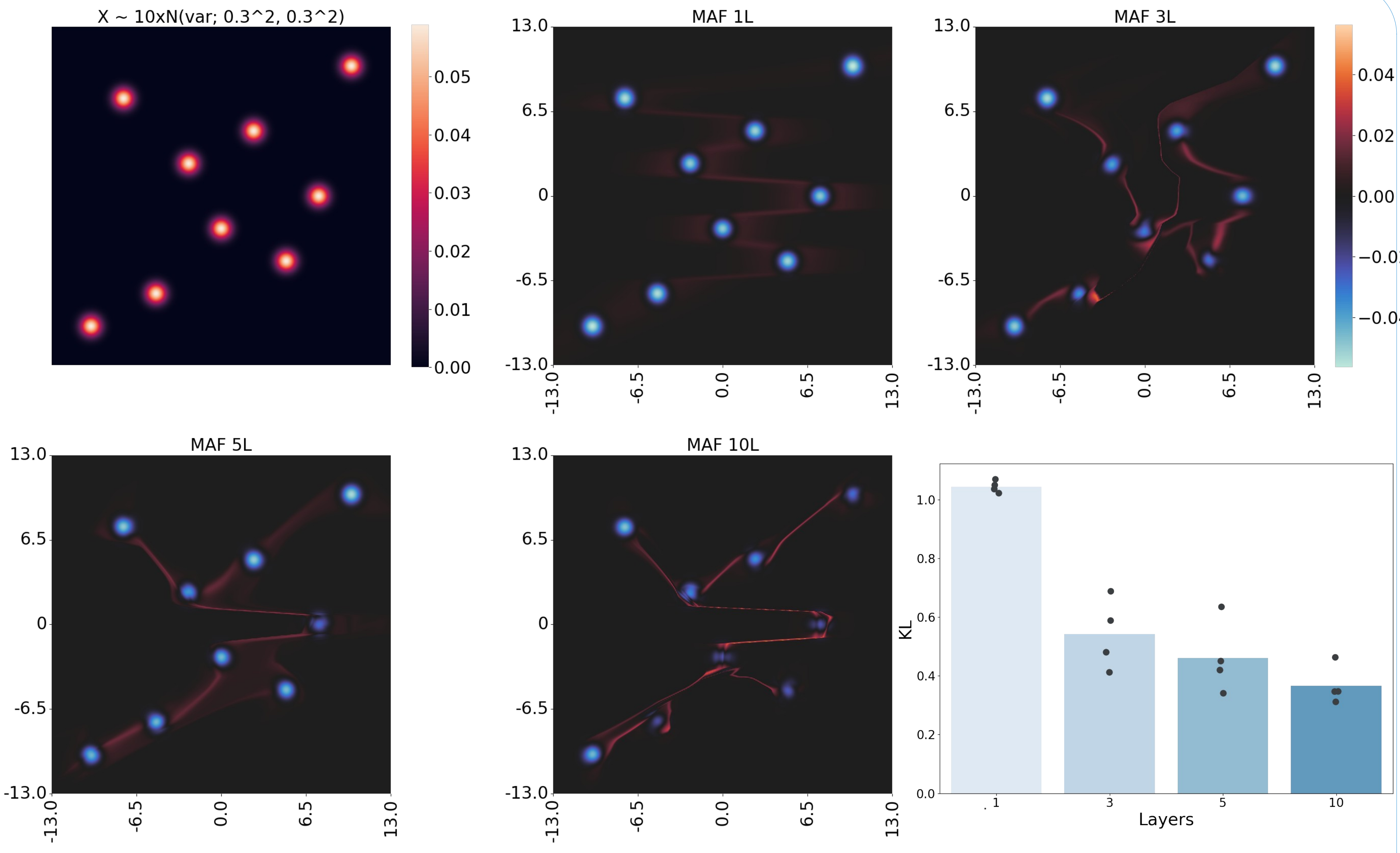
Density Estimation – 2D: 7 Gaussians



Density Estimation - 2D: 4 Gaussians



Density Estimation - 2D: 10 Gaussians



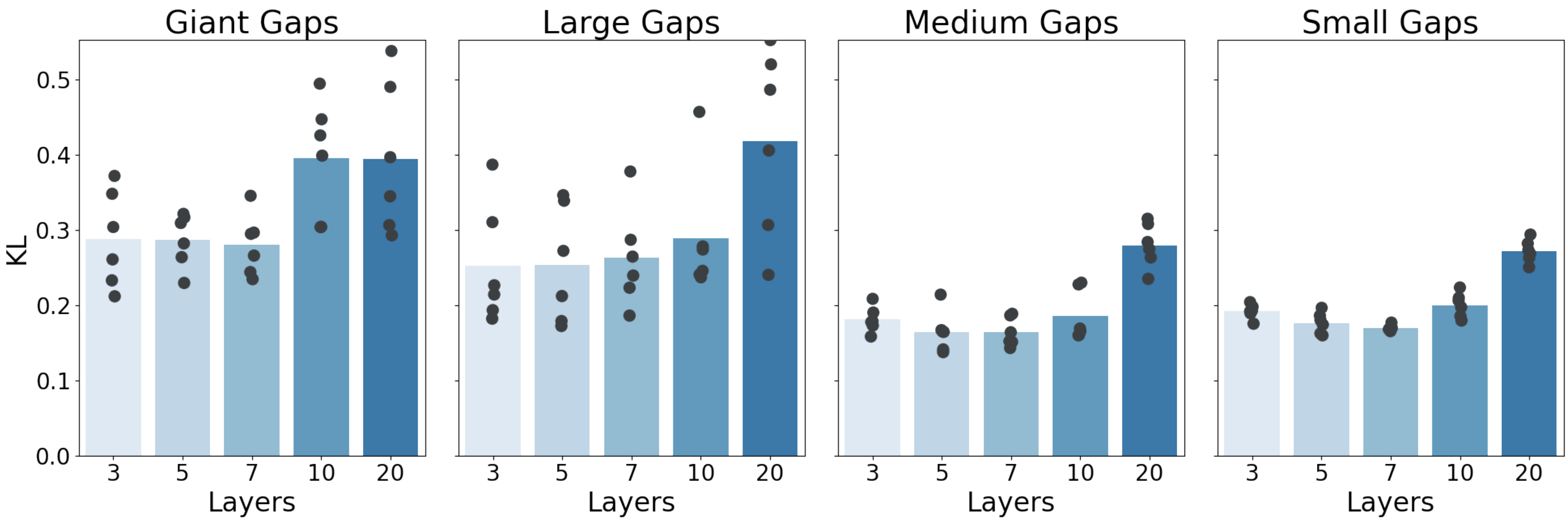
Density Estimation - 2D: KL Overview

	Layers: 1	3	5	10
10 Gaussians	1.04	0.54	0.46	0.37
2 Gaussians	0.32	0.03	0.03	0.02
3 Gaussians, row	0.74	0.22	0.04	0.02
3 Uniforms	0.89	0.13	0.11	0.08
4 Gaussians, diag	1.31	0.29	0.20	0.21
4 Gaussians, row	0.70	0.26	0.15	0.11
4 Uniforms	1.07	0.18	0.11	0.14
4 connected Gaussians	0.25	0.02	0.01	0.02
Simple Gaussian	0.00	0.00	0.00	0.00
Simple Uniform	0.20	0.04	0.02	0.01

Density Estimation - 2D: Summary

- Found:
  - More layers generally improve results
  - Layers >> Dimensions
  - ‘Gaps’ increase divergence
- Does this hold true for higher dimensions?

Density Estimation - 10D: Results and overview



Density Estimation - 10D: Summary

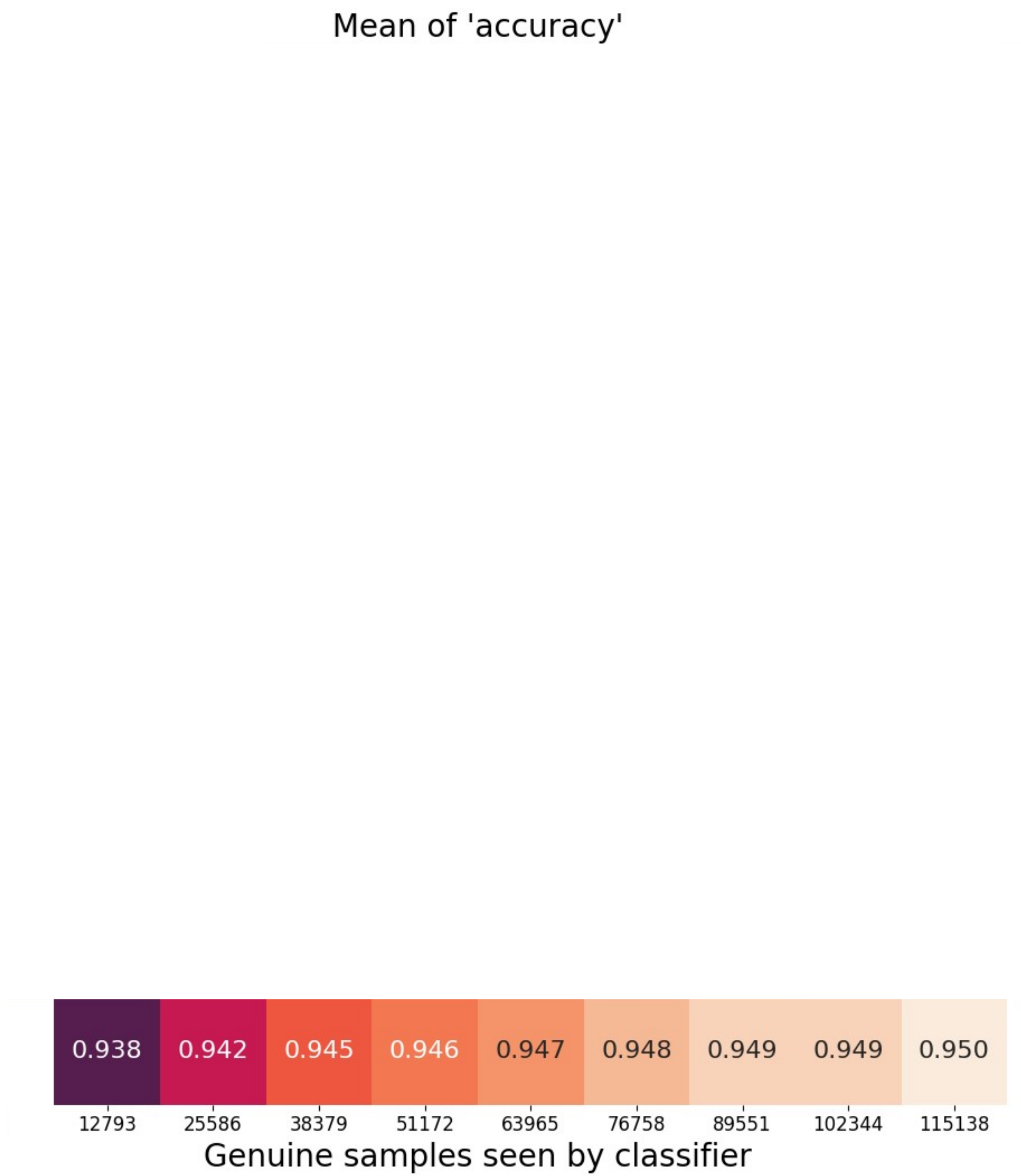
- Found:
  - More layers may make it worse
  - 5 to 7 layers works best
  - Layers < Dimensions
  - ‘Gaps’ increase divergence

Sampling Quality: MiniBooNE

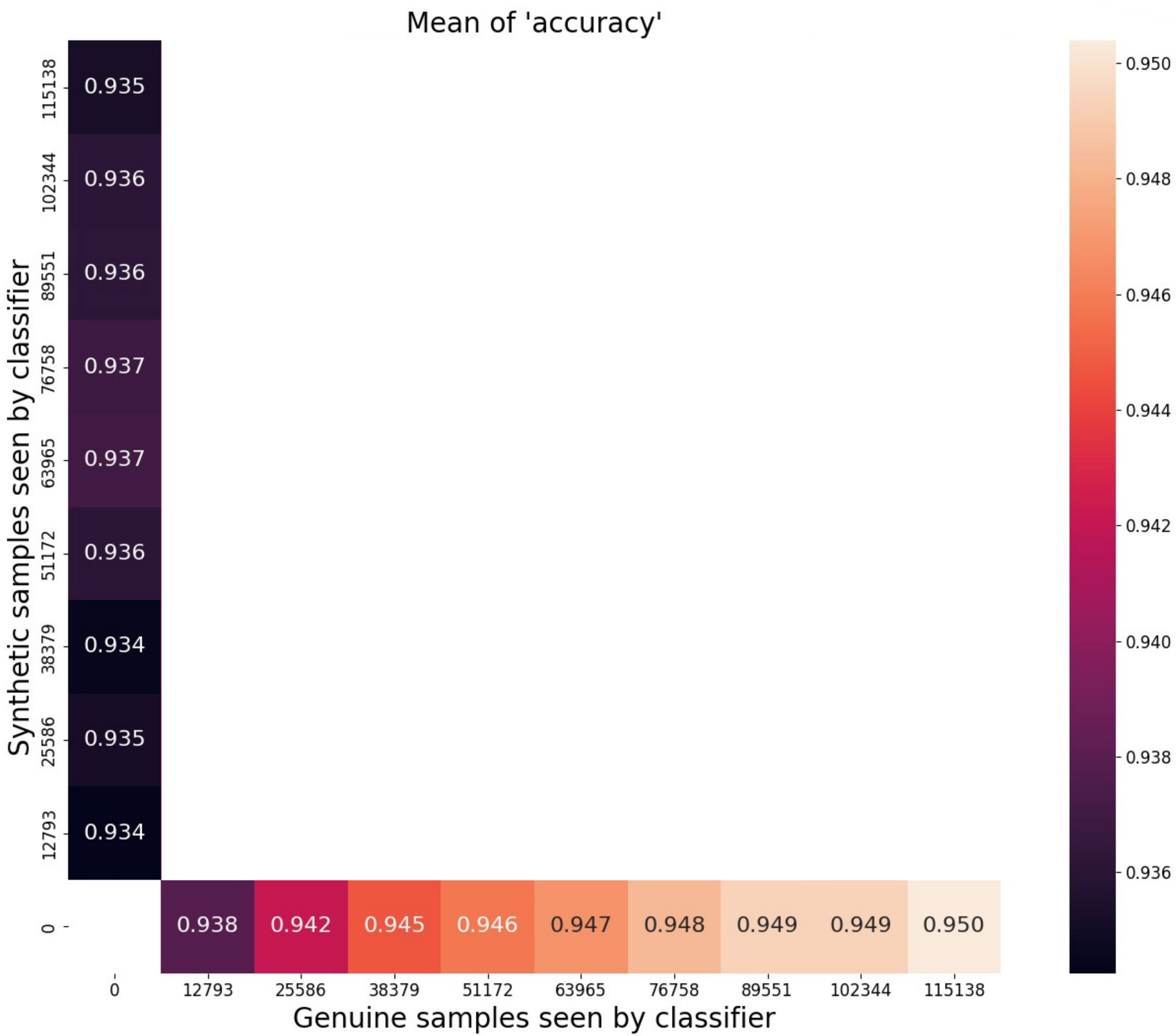
- 50 dimensions
- Physical experiment to observe neutrinos
- Binary classification problem: signal and noise
- ~130k samples
  - ~70% noise, ~30% signal
- Balanced and unbalanced experiment
- Experiment
  - Fit individual 30 layer MAFs for noise and signal
  - Sample synthetic dataset from MAFs
  - Fit classifiers with varying amounts of genuine and synthetic samples
  - Measure accuracy on test set

Universität Rostock Traditio et Innovatio

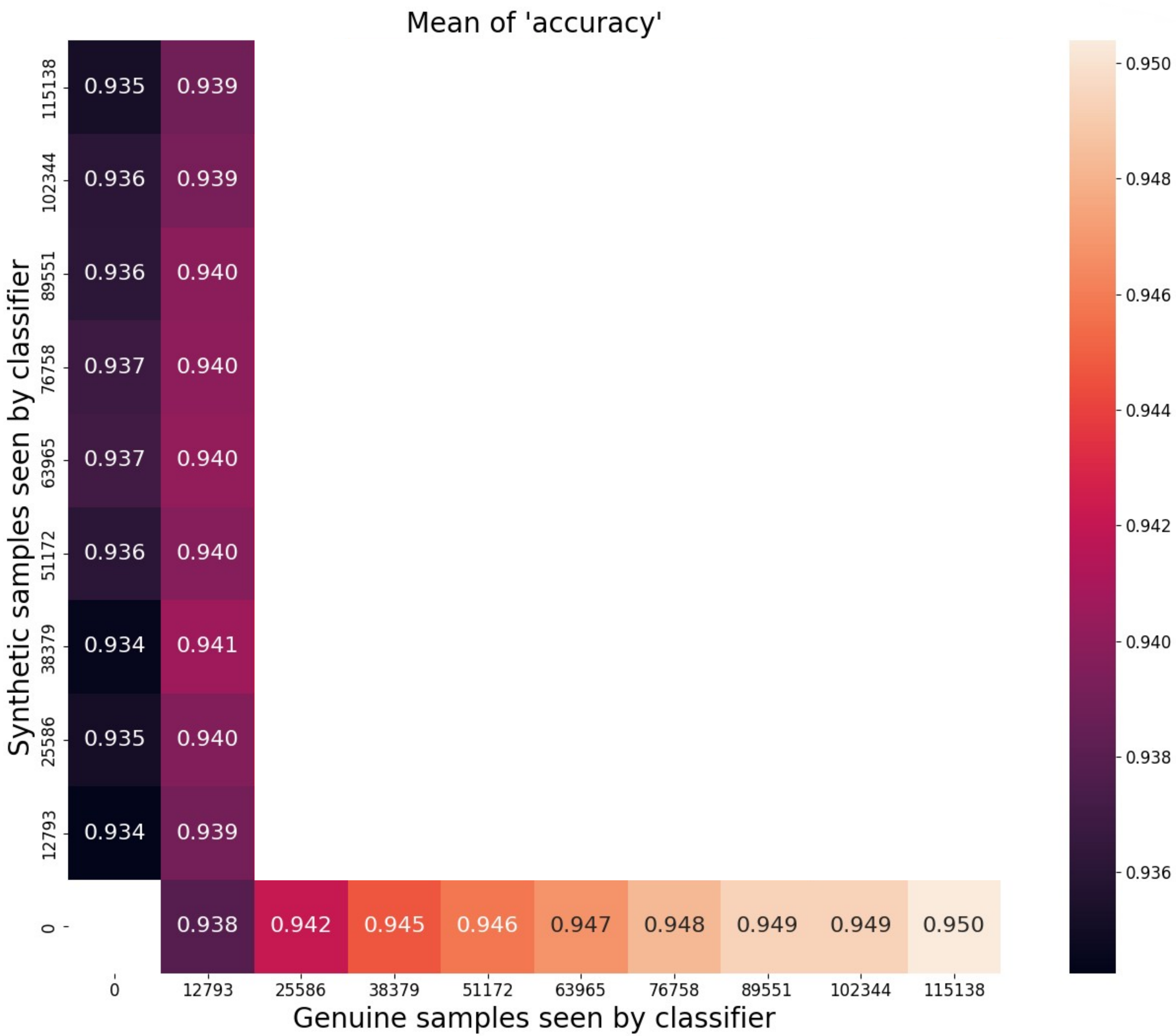
Sampling Quality: MiniBooNE Accuracy



Sampling Quality: MiniBooNE Accuracy

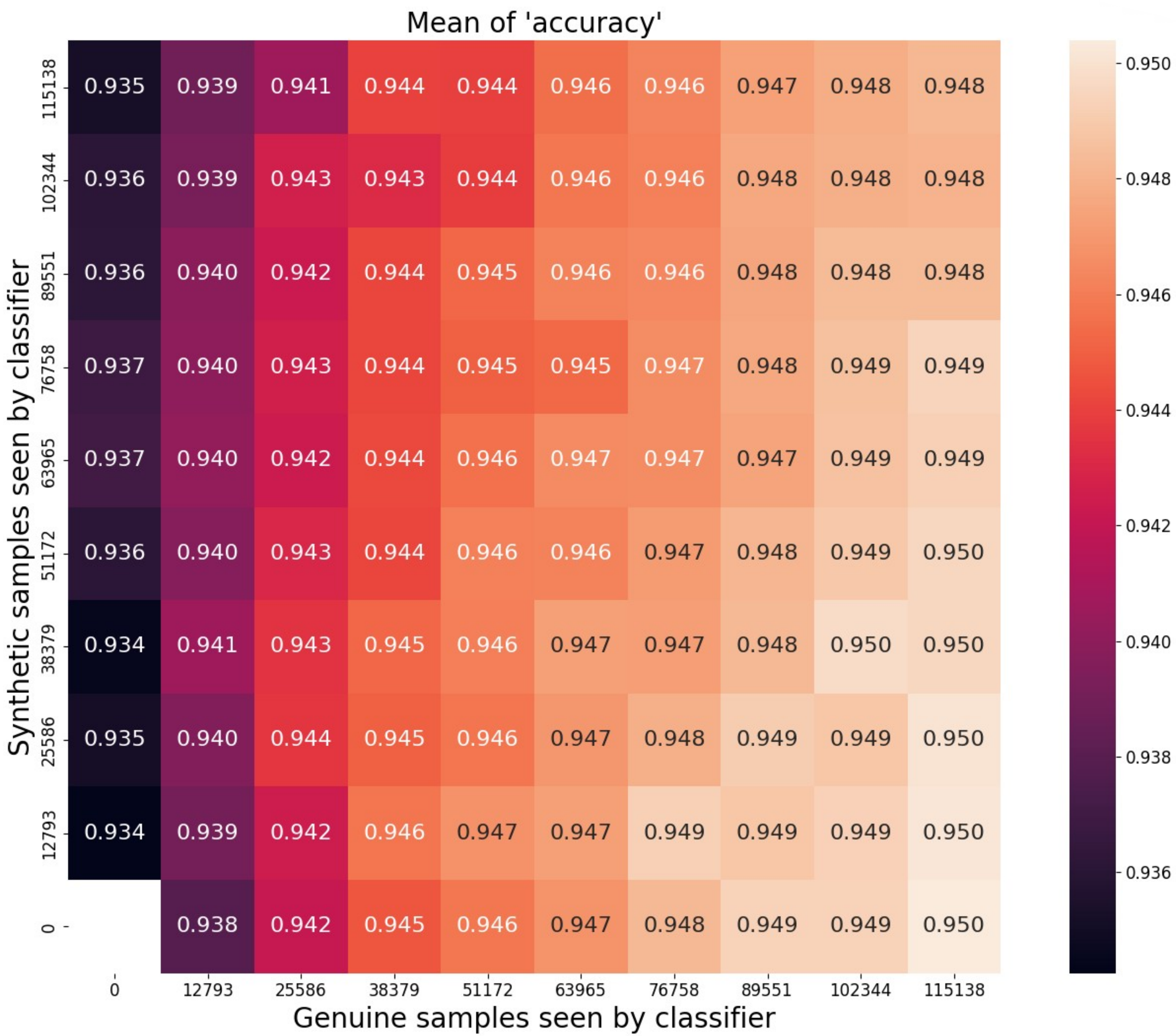


Sampling Quality: MiniBooNE Accuracy

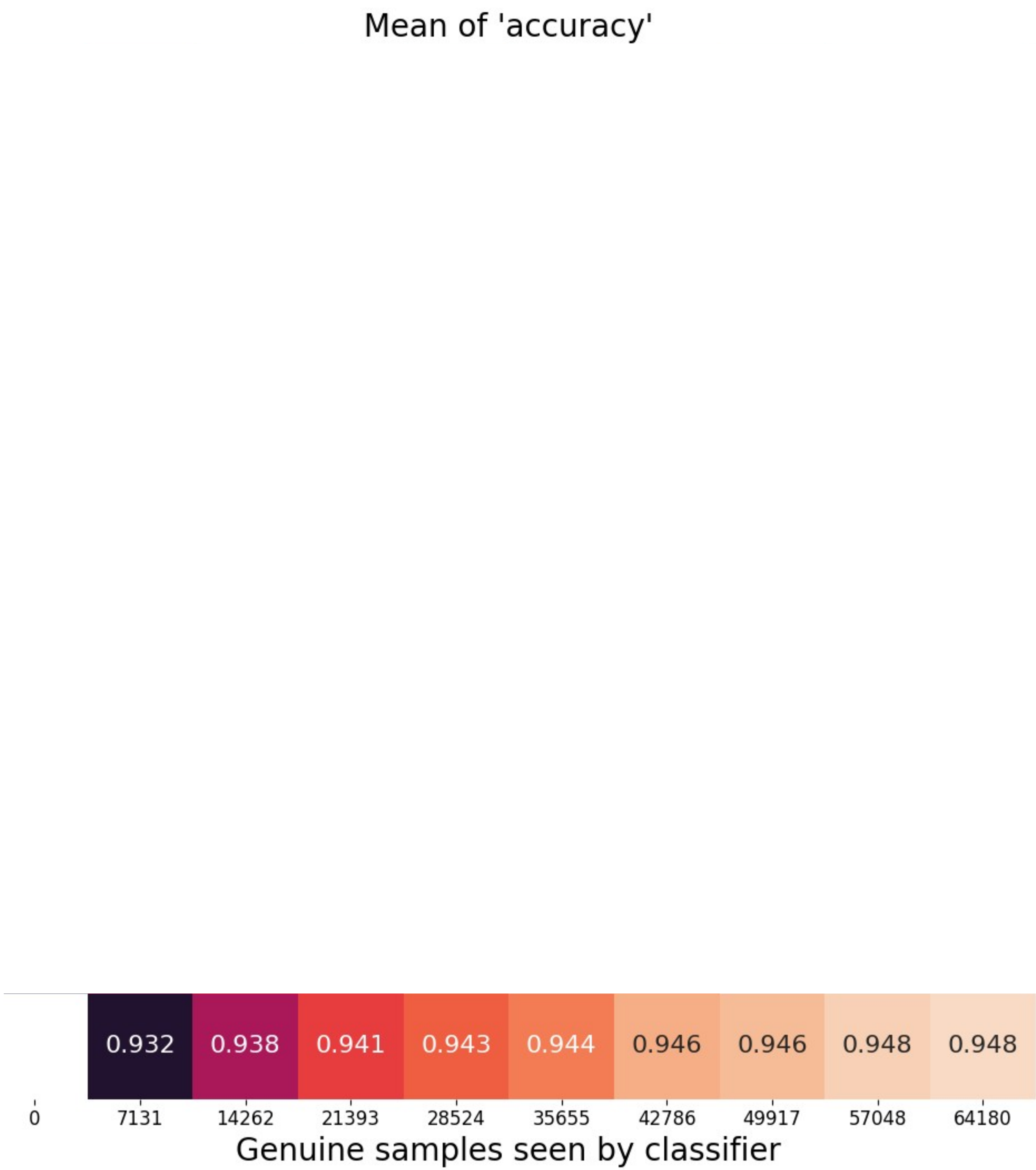


Universität Rostock Traditio et Innovatio

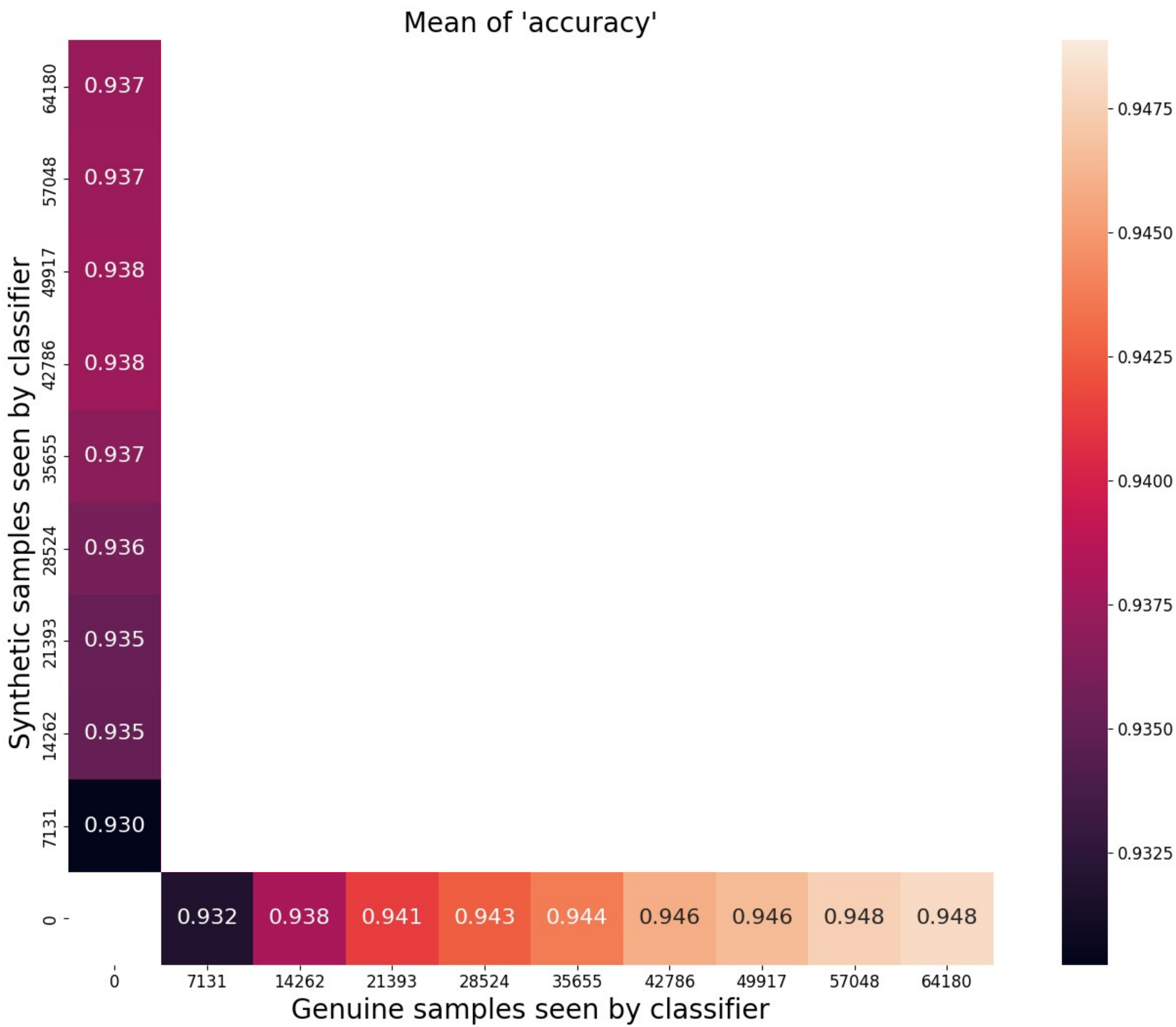
Sampling Quality: MiniBooNE Accuracy



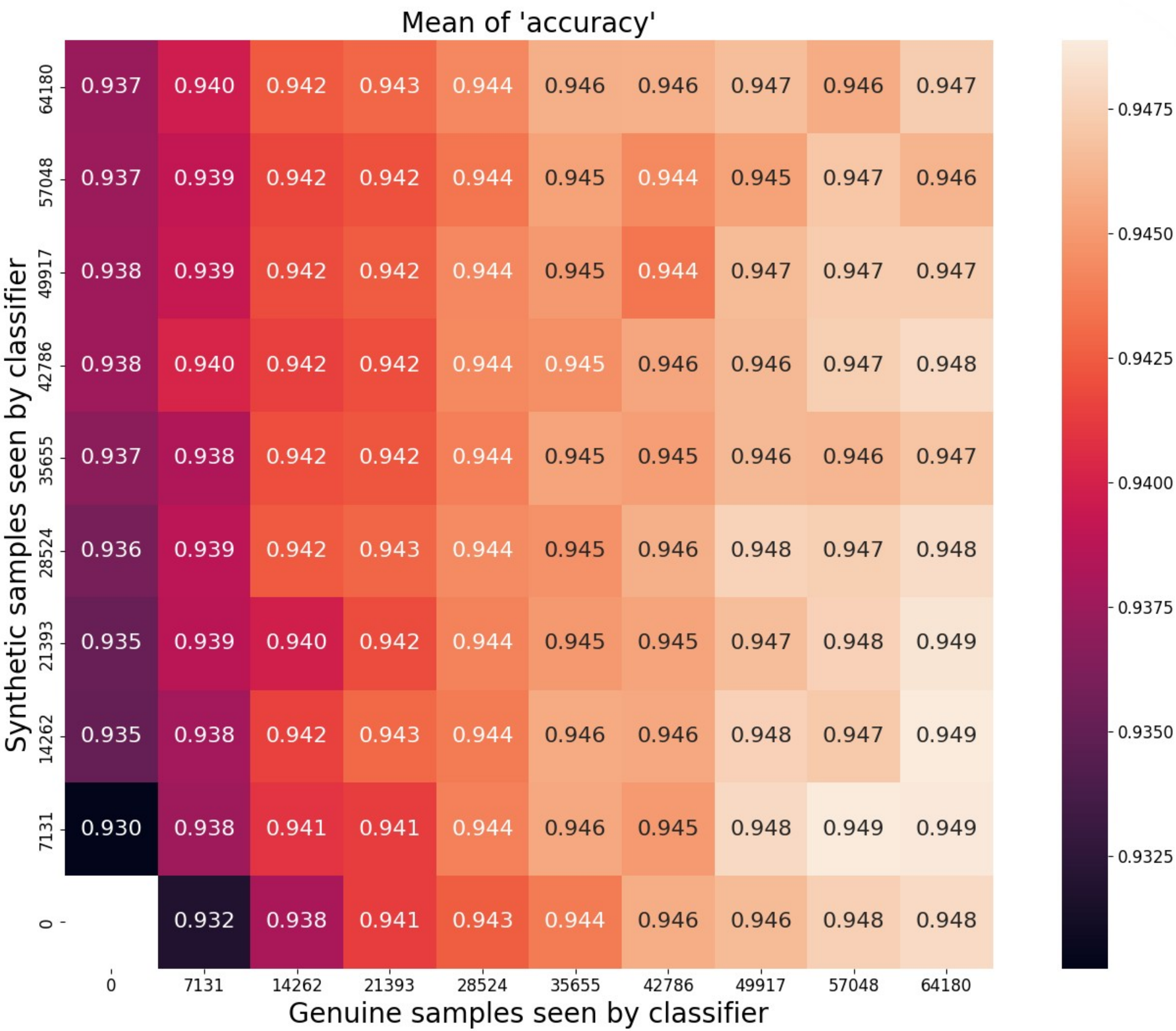
Sampling Quality: MiniBooNE, balanced



Sampling Quality: MiniBooNE, balanced



Sampling Quality: MiniBooNE, balanced



Sampling Quality: MiniBooNE Results

- Switching to synthetic samples
  - Accuracy drops slightly in all cases
- Augmenting synthetic samples
  - Complete dataset
    - Does no do anything at best
    - Augmenting many samples is harmful
  - Balanced
    - Few additional samples may improve accuracy very slightly
    - Augmenting many samples is harmful

Takeaway

- MAFs can learn arbitrary distributions
- MAFs are easy to tune
- MAFs scale well for higher dimensions

The End