UNIVERSITY OF CAMBRIDGE

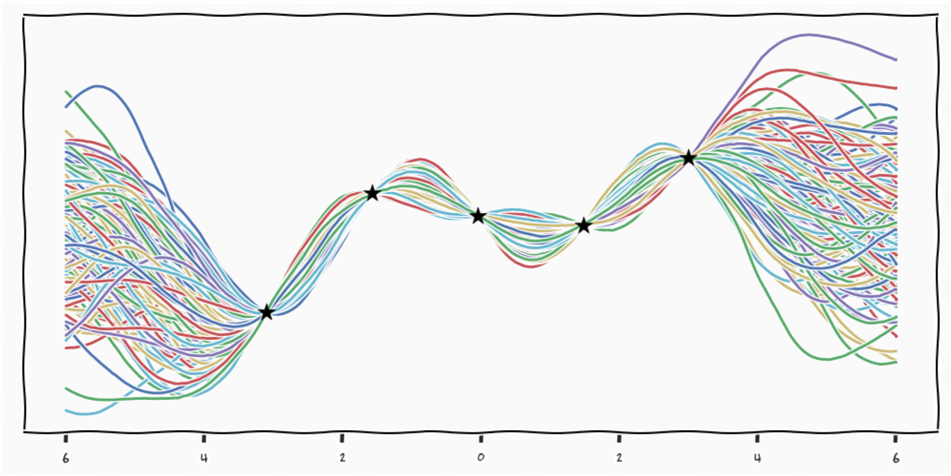
Machine Learning and the Physical World

Lecture 14 : Generative Models

Carl Henrik Ek - che29@cam.ac.uk

26th of November, 2025

http://carlhenrik.com


6
4
2
0
2
4
6

ChatGPT:
write an ode
to Ed Jaynes

Prophet of Probable Truths

O Jaynes, whose mind with reason burned,
Where others guessed, you deeply turned—
To principles, to logic tight,
You carved out paths through Bayesian light.

No mystic veil, no black-box creed,
But inference grounded in human need:
To reason with what we truly know,
And let the priors gently grow.

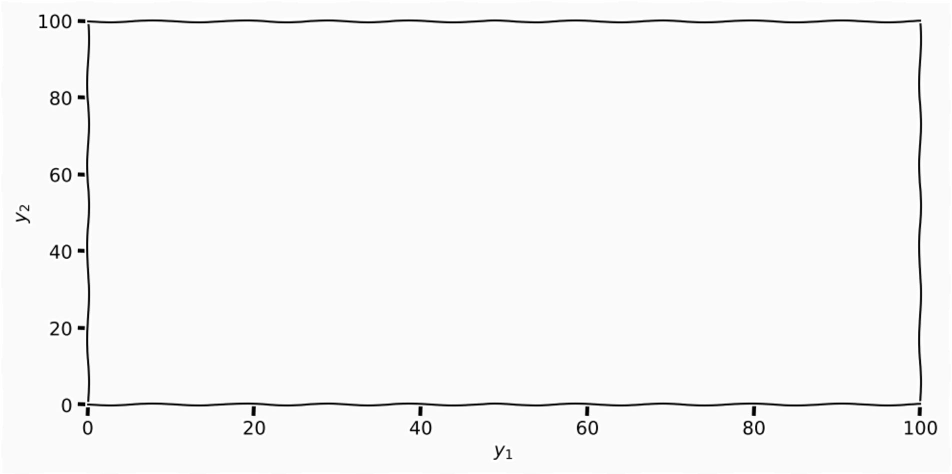
DeepAI: generate an image of a beaver eating a watermelon

β vs θ

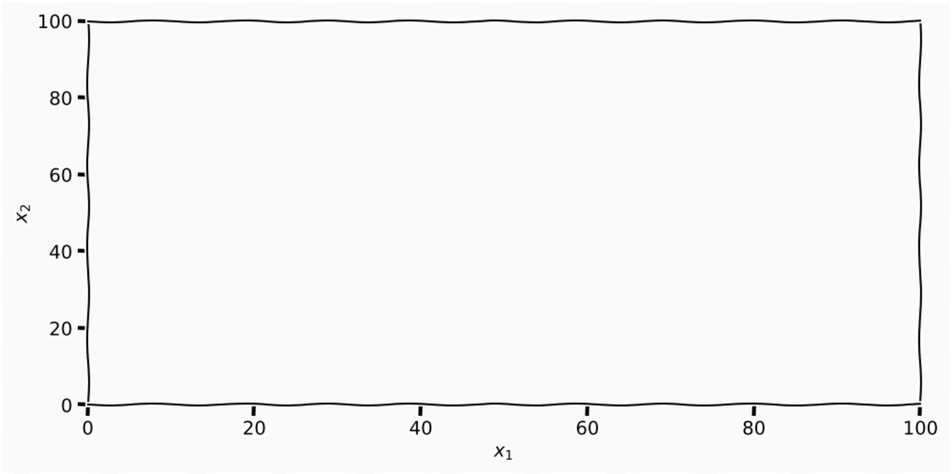
$$p(\theta \mid \mathcal{D}) = \frac{p(\mathcal{D} \mid \theta)p(\theta)}{p(\mathcal{D})}$$

$$p(y \mid \mathcal{D}) = \int p(y \mid \theta)p(\theta \mid \mathcal{D})\mathrm{d}\theta$$

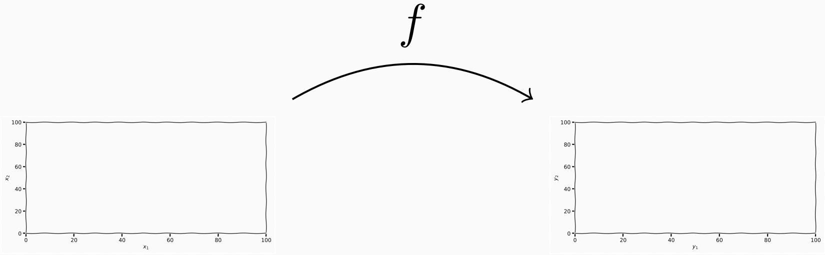
What is a Generative Model



Parameter Space

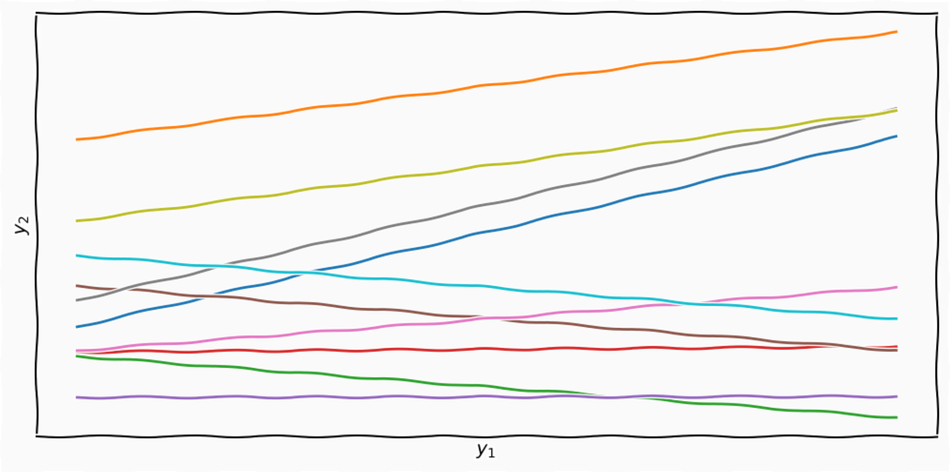


Generative Mapping



$$y = f(x)$$

Linear Regression

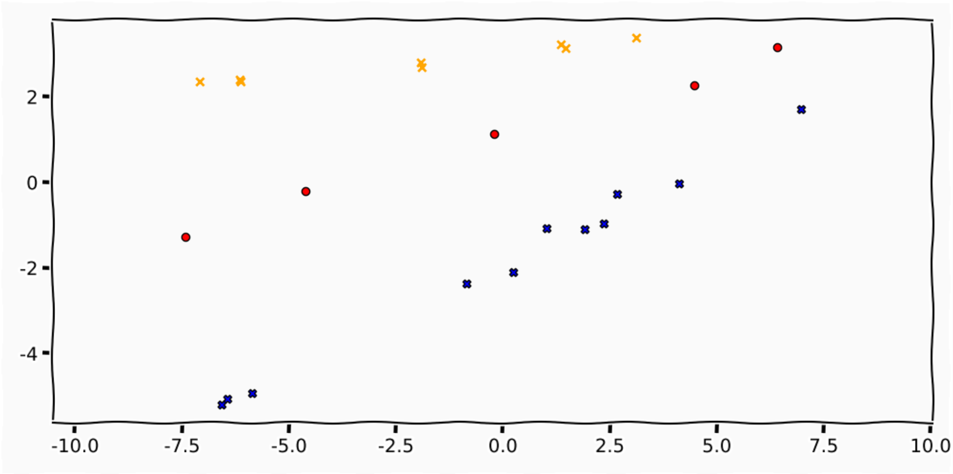


Generative Mapping

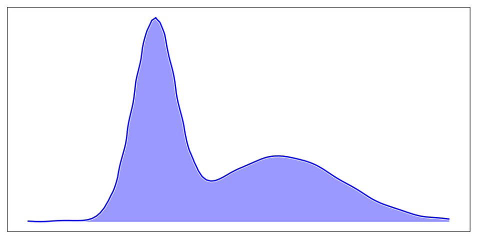
$$w_1, w_2 \xrightarrow{y = w_1 \cdot x + w_2} \texttt{ax.plot(x,y)}$$

Data is computations that the universe have already done for us
– Prof. Neil D. Lawrence

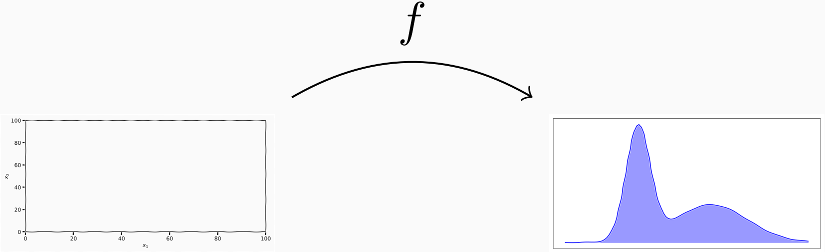
Linear Regression with Data



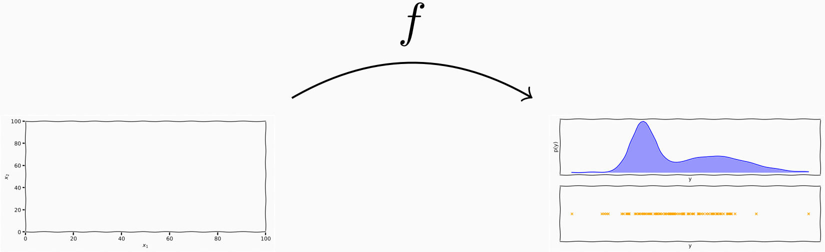
Data



Generative Mapping



Generative Mapping



Generative Models

- Where should we introduce uncertainty?

Generative Models

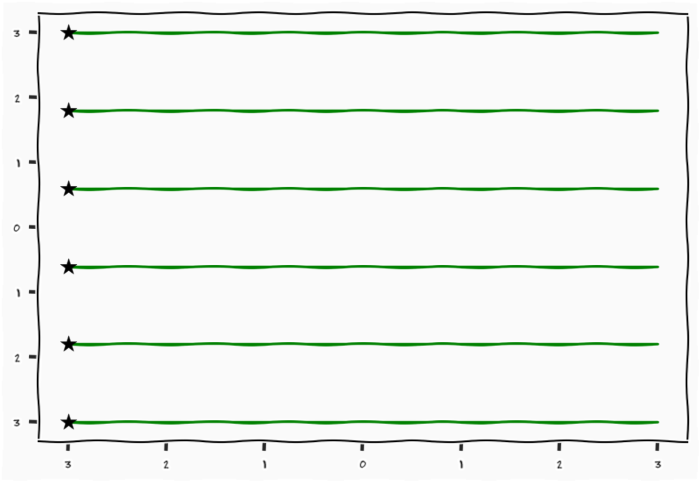
- Where should we introduce uncertainty?
- How should we choose the generative function?

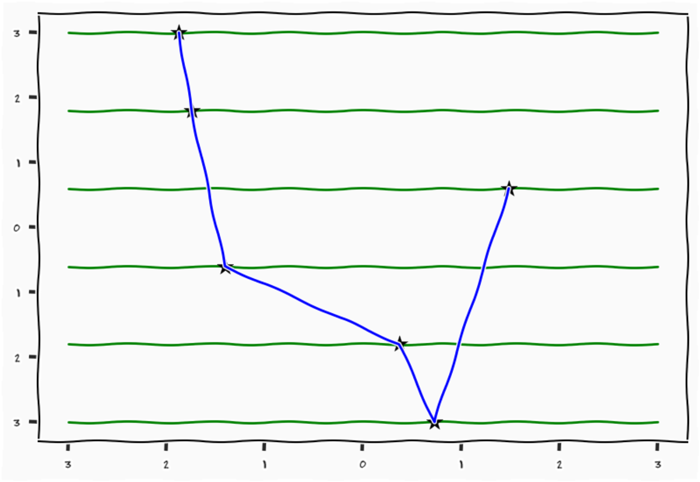
Generative Models

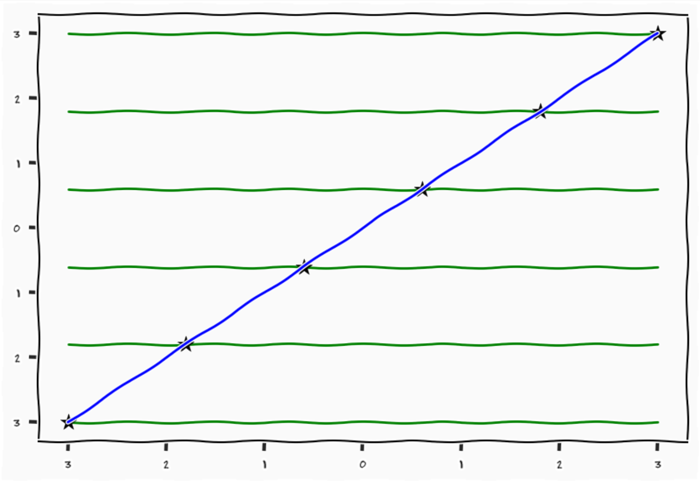
- Where should we introduce uncertainty?
- How should we choose the generative function?
- How should we choose latent representation?

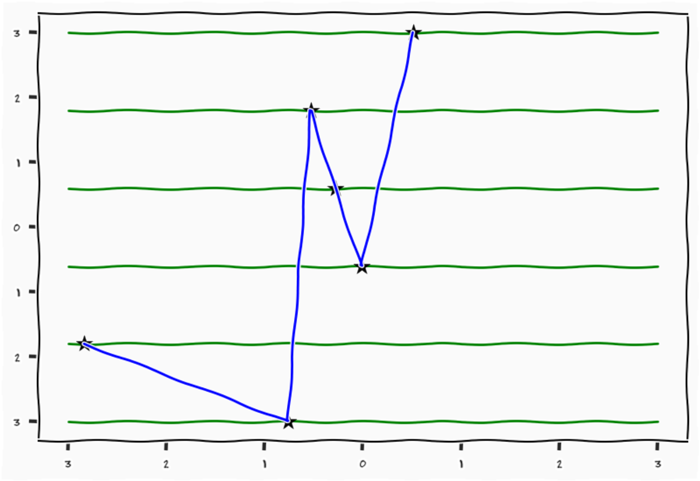
Generative Models

- Where should we introduce uncertainty?
- How should we choose the generative function?
- How should we choose latent representation?
- *How can we constrain the above using data?*


3
2
1
0
1
2
3
3
2
1
0
1
2
3

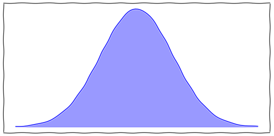

3
2
1
0
1
2
3
3
2
1
0
1
2
3


3
2
1
0
1
2
3
3
2
1
0
1
2
3

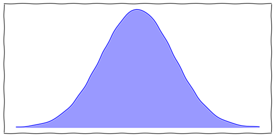

3
2
1
0
1
2
3
3
2
1
0
1
2
3

PCA “Probabilistic principal component analysis”

$$p(x)$$



$$\int p(y \mid x) p(x) \mathrm{d}x$$

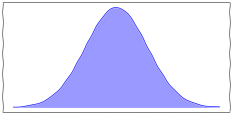


$$\mathbf{w}$$

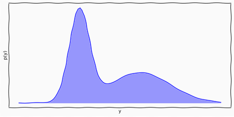
$$p(y)$$

GP-LVM “Gaussian Process Models for Visualisation of High Dimensional Data”

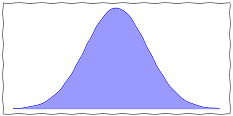
$$p(x)$$



$$\int p(y \mid f)p(f \mid x)p(x)\mathrm{d}x\mathrm{d}f$$

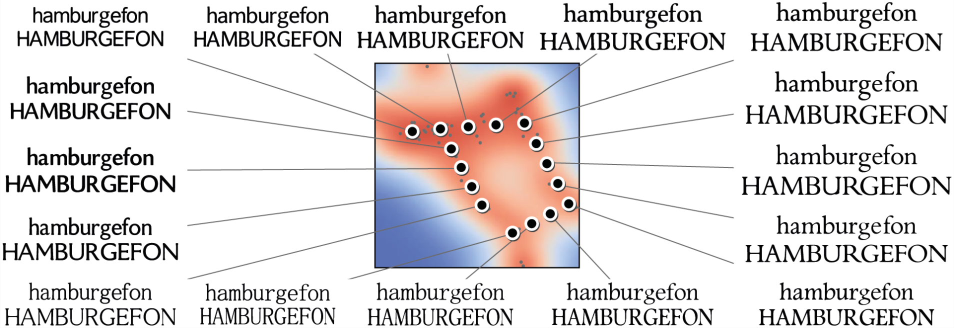


$$p(y)$$

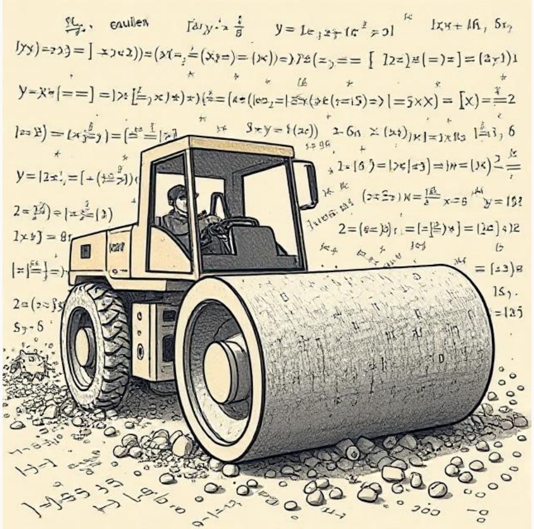


$$f \sim \mathcal{GP}$$

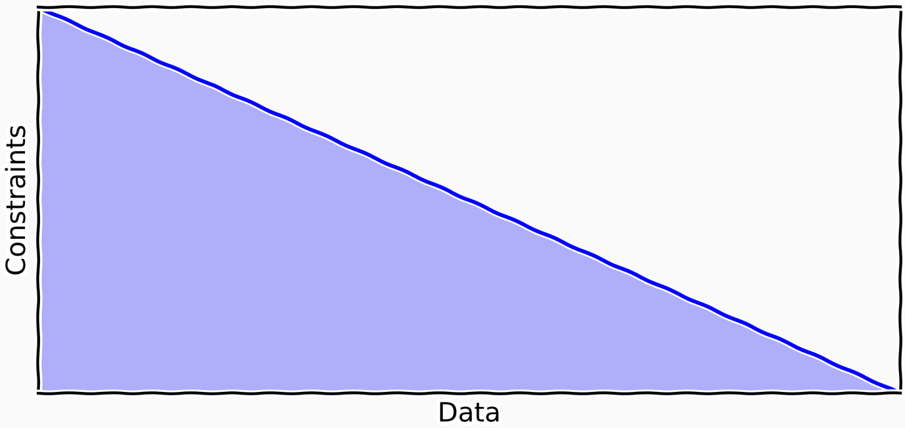
Fonts "Learning a manifold of fonts"



URL



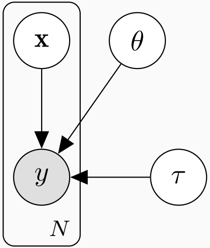
Data vs Constraints



November 26, 2025

Regularisation

Variational Auto-encoders “Auto-Encoding Variational Bayes”



$$p(y) = \int \prod_i^N p(y_i \mid x_i) p(x_i) \mathrm{d}x_i$$

$$\mathrm{KL}(q(x \mid y) \mid\mid p(x \mid y)) = \int q(x \mid y) \log \frac{q(x \mid y)}{p(x \mid y)} \mathrm{d}x$$

Variational Auto-encoders "Auto-Encoding Variational Bayes"

$$\begin{aligned}
\mathrm{KL}(q(x \mid y) \mid\mid p(x \mid y)) &= \int q(x \mid y) \log \frac{q(x \mid y)}{p(x \mid y)} \mathrm{d}x \\
&= \int q(x \mid y) \log \frac{q(x \mid y)}{\left(\frac{p(y|x)p(x)}{p(y)}\right)} \mathrm{d}x
\end{aligned}$$

Variational Auto-encoders “Auto-Encoding Variational Bayes”

$$
\begin{aligned}
\mathrm{KL}(q(x \mid y) \mid\mid p(x \mid y)) &= \int q(x \mid y) \log \frac{q(x \mid y)}{p(x \mid y)} \mathrm{d}x \\
&= \int q(x \mid y) \log \frac{q(x \mid y)}{\left(\frac{p(y|x)p(x)}{p(y)}\right)} \mathrm{d}x \\
&= \log p(y) + \int q(x \mid y) \log \frac{q(x \mid y)}{p(x)} \mathrm{d}x - \int q(x \mid y) \log p(y \mid x) \mathrm{d}x
\end{aligned}
$$

Variational Auto-encoders “Auto-Encoding Variational Bayes”

$$\mathrm{KL}(q(x \mid y) \mid\mid p(x \mid y)) = \int q(x \mid y) \log \frac{q(x \mid y)}{p(x \mid y)} \mathrm{d}x$$

$$= \int q(x \mid y) \log \frac{q(x \mid y)}{\left(\frac{p(y|x)p(x)}{p(y)}\right)} \mathrm{d}x$$

$$= \log p(y) + \int q(x \mid y) \log \frac{q(x \mid y)}{p(x)} \mathrm{d}x - \int q(x \mid y) \log p(y \mid x) \mathrm{d}x$$

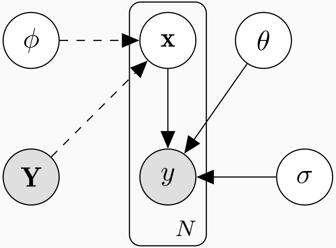
$$= \log p(y) + \mathrm{KL}(q(x \mid y) \mid\mid p(x)) - \int q(x \mid y) \log p(y \mid x) \mathrm{d}x$$

$$
\begin{aligned}
\log p(y) &= \int q(x \mid y) \log p(y \mid x) \mathrm{d}x - \mathrm{KL}(q(x \mid y) \mid\mid p(x)) \\
&\quad + \mathrm{KL}(q(x \mid y) \mid\mid p(x \mid y)) \\
&\geq \int q(x \mid y) \log p(y \mid x) \mathrm{d}x - \mathrm{KL}(q(x \mid y) \mid\mid p(x))
\end{aligned}
$$

VAE Objective

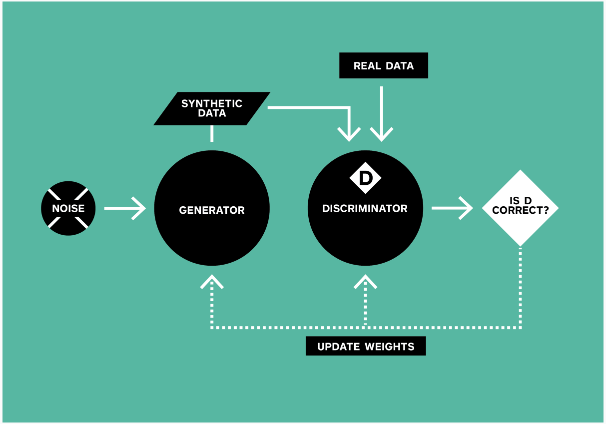
$$
\hat{\theta}, \hat{\phi} = \mathrm{argmax}_{\theta,\phi} \, \mathbb{E}_{q_\phi(x|y)} \left[\log p_\theta(y \mid x)\right] - \mathrm{KL}(q_\phi(x \mid y) \mid\mid p(x))
$$

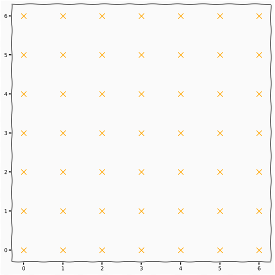
Variational Auto-encoders “Auto-Encoding Variational Bayes”

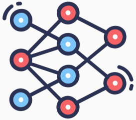


$$\hat{\theta}, \hat{\phi} = \mathrm{argmax}_{\theta,\phi}\, \mathbb{E}_{q_\phi(x|y)} \left[\log p_\theta(y \mid x)\right] - \mathrm{KL}(q_\phi(x \mid y) \mid\mid p(x))$$

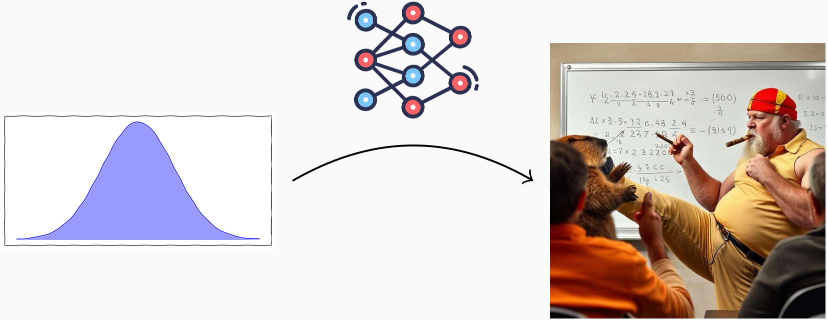
Generative Adversarial Networks “A Possibility for Implementing Curiosity and Boredom in Model-Building Neural Controllers”




0
1
2
3
4
5
6



All about the function



Integration by Substitution

$$\int_a^b f(x)\mathrm{d}x = \int_{y(a)}^{y(b)} f(x(y))\frac{\mathrm{d}x}{\mathrm{d}y}\mathrm{d}y$$

Transformation function $y : \mathcal{X} \to \mathcal{Y}$

Inverse $x : \mathcal{Y} \to \mathcal{X}$

Transformation of Densities

$$P(a \leq x < b) = \int_a^b p(x)\mathrm{d}x$$

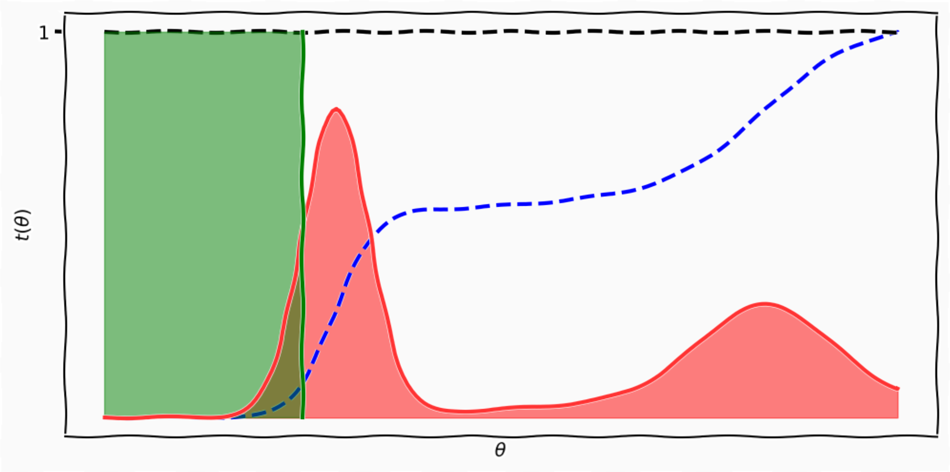
$$P(y(a) \leq y < y(b)) = \int_{y(a)}^{y(b)} p(x(y))\frac{\mathrm{d}x}{\mathrm{d}y}\mathrm{d}y$$

$$\Rightarrow p(y) = p(x(y))\frac{\mathrm{d}x}{\mathrm{d}y}$$

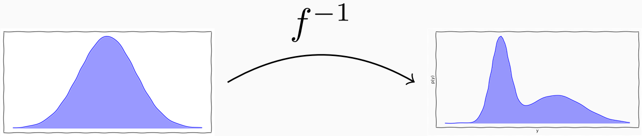
$p(x)$ is a density function

$$p(y) = p(x(y))\left|\frac{\mathrm{d}x}{\mathrm{d}y}\right|$$

Change of Variable

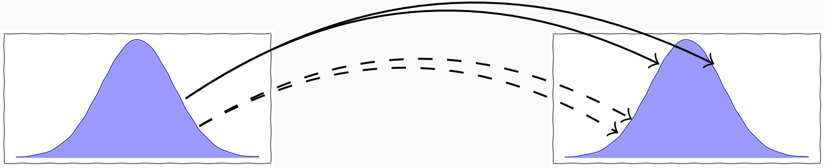


Normalising Flows "Normalizing Flows: an Introduction and Review of Current Methods"

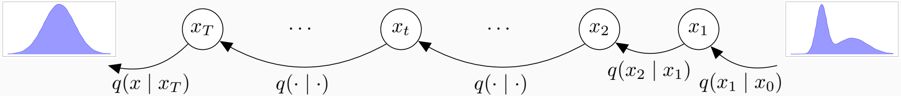


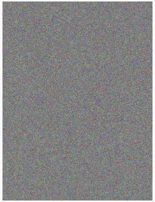
$$\operatorname{argmax}_\theta \log(p(y)) = \log\left(\left|\frac{\mathrm{d}f_\theta}{\mathrm{d}y}\right|\right) + \log(p(f_\theta(y)))$$

Equivalence Class



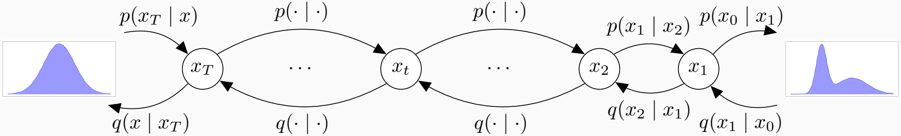

x_T
$\cdots$
x_t
$\cdots$
x_2
x_1

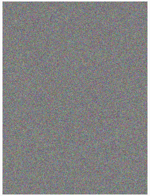

x_T
$\cdots$
x_t
$\cdots$
x_2
x_1
$q(x \mid x_T)$
$q(\cdot \mid \cdot)$
$q(\cdot \mid \cdot)$
$q(x_2 \mid x_1)$
$q(x_1 \mid x_0)$



$\mathcal{N}(\mathbf{x}_T; \boldsymbol{\mu}, \boldsymbol{\Sigma})$

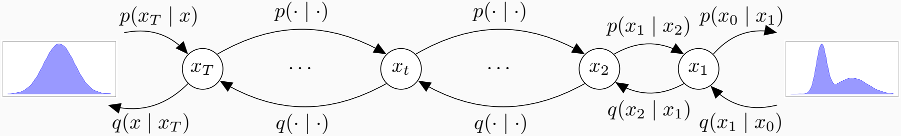
$\mathbf{x_0}$


$p(x_T \mid x)$
$q(x \mid x_T)$
x_T
$p(\cdot \mid \cdot)$
$\cdots$
$q(\cdot \mid \cdot)$
x_t
$p(\cdot \mid \cdot)$
$\cdots$
$q(\cdot \mid \cdot)$
x_2
$p(x_1 \mid x_2)$
$q(x_2 \mid x_1)$
x_1
$p(x_0 \mid x_1)$
$q(x_1 \mid x_0)$

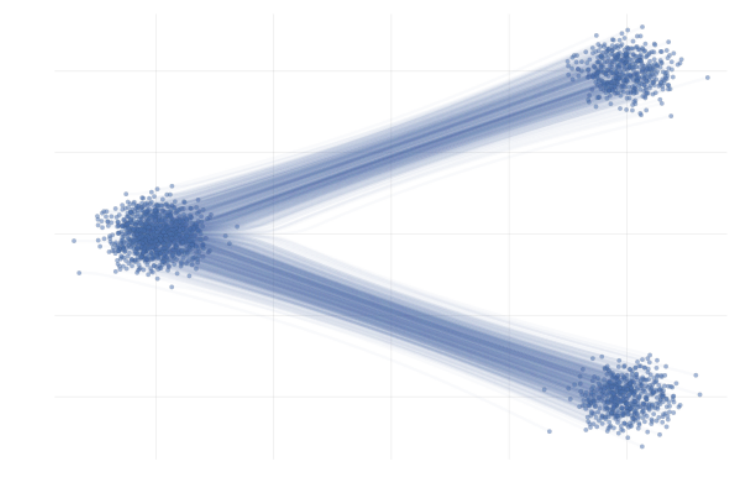


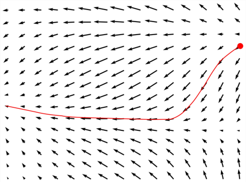
$$\mathcal{N}(\mathbf{x}_T; \boldsymbol{\mu}, \boldsymbol{\Sigma})$$

$$p(\mathbf{x}_0 \mid \mathbf{x}_T)$$


$p(x_T \mid x)$
$q(x \mid x_T)$
x_T
$p(\cdot \mid \cdot)$
$\cdots$
$q(\cdot \mid \cdot)$
x_t
$p(\cdot \mid \cdot)$
$\cdots$
$q(\cdot \mid \cdot)$
x_2
$p(x_1 \mid x_2)$
$q(x_2 \mid x_1)$
x_1
$p(x_0 \mid x_1)$
$q(x_1 \mid x_0)$

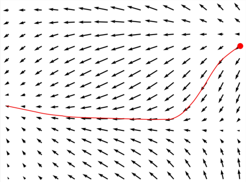
Flow Matching





- Ordinary Differential Equation

$$X : [0,1] \to \mathbb{R}^d, \quad t \to X_t$$

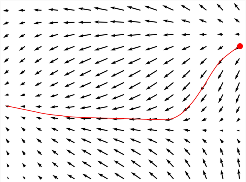


- Ordinary Differential Equation

$$X : [0, 1] \to \mathbb{R}^d, \quad t \to X_t$$

- Vector Field (defines ODE)

$$u : \mathbb{R}^d \times [0, 1] \to \mathbb{R}^d, \quad (x, t) \to u_t(x)$$



- Ordinary Differential Equation

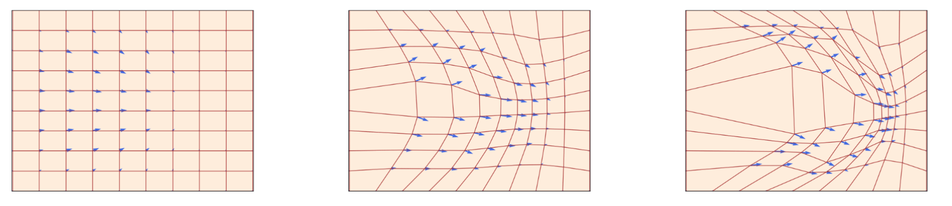
$$X : [0,1] \to \mathbb{R}^d, \quad t \to X_t$$

- Vector Field (defines ODE)

$$u : \mathbb{R}^d \times [0,1] \to \mathbb{R}^d, \quad (x,t) \to u_t(x)$$

- Flow

$$\psi : \mathbb{R}^d \times [0,1] \to \mathbb{R}^d$$
$$\frac{\mathrm{d}}{\mathrm{d}t}\psi_t(x_0) = u_t(\psi(x_0))$$



Existence
If u is continuously differentiable with a bounded derivative, then the ODE in has a unique solution given by a flow ψ_t. In this case the flow is a *diffeomorphism* for all t.

Taking Stock II

- Normalising flows - Density transformation

Taking Stock II

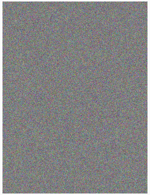
- Normalising flows - Density transformation
- Diffusion Models - Stochastic Differential Equation

Taking Stock II

- Normalising flows - Density transformation
- Diffusion Models - Stochastic Differential Equation
- Flow Matching - Ordinary Differential Equation

Diffusion Models as Simulators

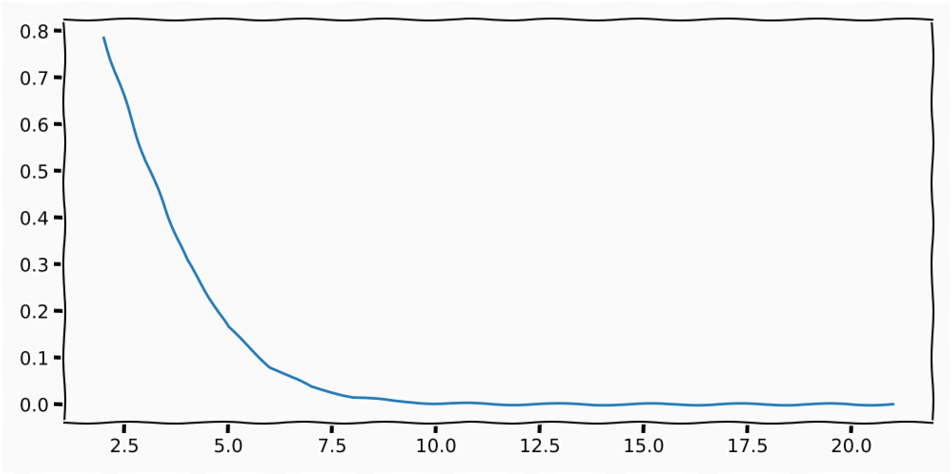
Diffusion Models

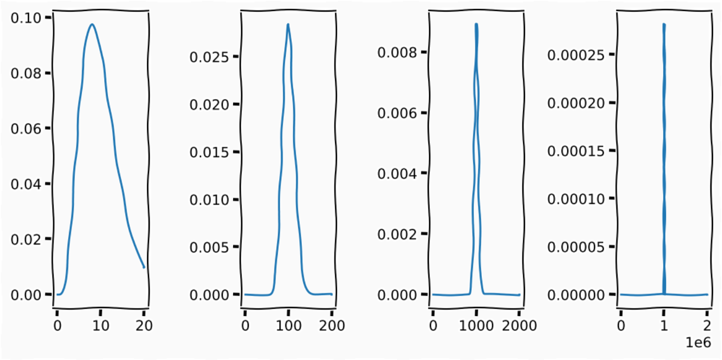


$$\mathcal{N}(\mathbf{x}_T; \boldsymbol{\mu}, \boldsymbol{\Sigma})$$

$$p(\mathbf{x}_0 \mid \mathbf{x}_T)$$

Hypercube vs Hypersphere

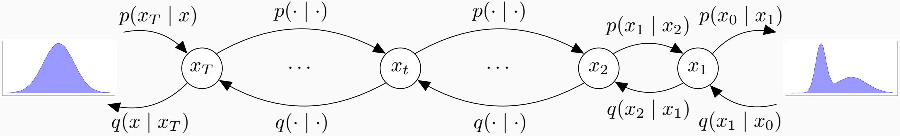




Theorem (Gaussian Annullus)

$$\|\mathcal{N}(\mathbf{0}, \mathbf{I}_{d\times d})\|^2 \sim \chi^2_d$$

Diffusion Model



Diffusion Models Characteristics

- Excellent parametrisation of support of complex distributions

Diffusion Models Characteristics

- Excellent parametrisation of support of complex distributions
- They are very high-dimensional

Diffusion Models Characteristics

- Excellent parametrisation of support of complex distributions
- They are very high-dimensional
- They are not density models but based on samples

Diffusion Models Characteristics

- Excellent parametrisation of support of complex distributions
- They are very high-dimensional
- They are not density models but based on samples
- Challenging to control support of generation

Task

- Parametrisation the support of the model

Task

- Parametrisation the support of the model
- Low dimensional subspace representation that allows for emulation

Task

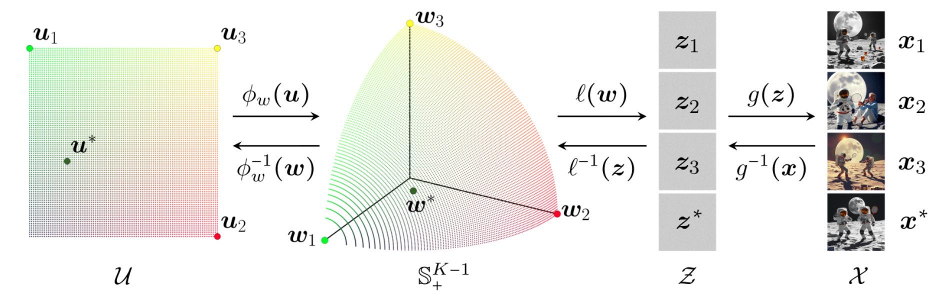
- Parametrisation the support of the model
- Low dimensional subspace representation that allows for emulation
- Non-parametric subspace

Task

- Parametrisation the support of the model
- Low dimensional subspace representation that allows for emulation
- Non-parametric subspace
- Bijective relationship between representation and model

Erik Bodin et al. (2024). "Linear Combinations of Gaussian Latents in Generative Models: Interpolation and Beyond". In: *Proceedings of the Thirteenth International Conference on Learning Representations (ICLR)*

Samuel Willis et al. (2025). "Define Latent Spaces By Example: Optimisation Over the Outputs of Generative Models". In: *CoRR*



u_1
u_3
u^*
u_2
$\mathcal{U}$
$\phi_w(u)$
$\phi_w^{-1}(w)$
w_3
w^*
w_1
w_2
$\mathbb{S}_+^{K-1}$
$\ell(w)$
$\ell^{-1}(z)$
z_1
z_2
z_3
z^*
$\mathcal{Z}$
$g(z)$
$g^{-1}(x)$
x_1
x_2
x_3
x^*
$\mathcal{X}$

$$\tilde{\mathbf{x}} = \sum_{k=1}^{K} w_k \mathbf{x}_k = \mathbf{w}^{\mathrm{T}} \mathbf{X}$$

Latent Space Manipulations

$$\tilde{\mathbf{x}} = \sum_{k=1}^{K} w_k \mathbf{x}_k = \mathbf{w}^{\mathrm{T}} \mathbf{X}$$

$$\tilde{\mathbf{x}} \sim \mathcal{N}\left(\sum_{k=1}^{K} w_k \boldsymbol{\mu}, \mathbf{w}^{\mathrm{T}} \mathbf{w} \boldsymbol{\Sigma} \right)$$

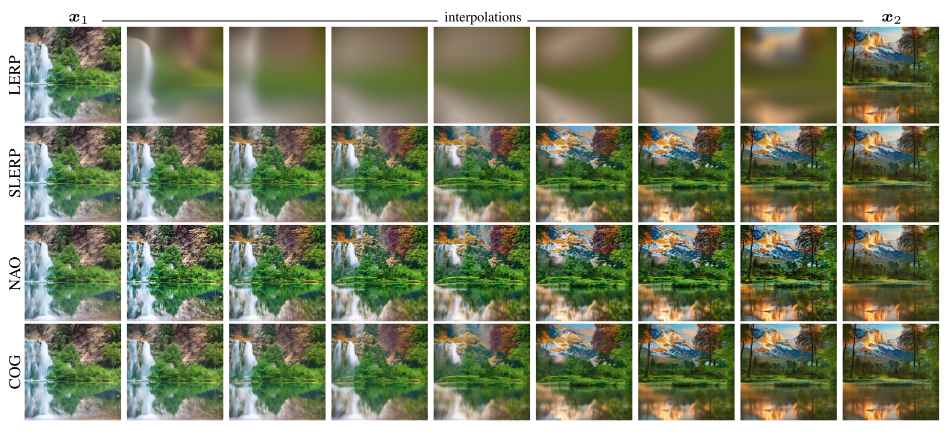
Latent Space Manipulations

$$\tilde{\mathbf{x}} = \sum_{k=1}^{K} w_k \mathbf{x}_k = \mathbf{w}^{\mathrm{T}} \mathbf{X}$$

$$\tilde{\mathbf{x}} \sim \mathcal{N}\left(\sum_{k=1}^{K} w_k \boldsymbol{\mu}, \mathbf{w}^{\mathrm{T}} \mathbf{w} \boldsymbol{\Sigma} \right)$$

$$\mathbf{z} = \mathcal{T}\left(\mathbf{w}^{\mathrm{T}} \mathbf{X}\right) = \left(1 - \frac{\sum_{k=1}^{K} w_k}{\sqrt{\mathbf{w}^{\mathrm{T}} \mathbf{w}}} \right) \boldsymbol{\mu} + \frac{\mathbf{w}^{\mathrm{T}} \mathbf{X}}{\mathbf{w}^{\mathrm{T}} \mathbf{w}}$$

Interpolation



Subspace


$\boldsymbol{x}_1$
$\boldsymbol{x}_2$
$\boldsymbol{x}_3$
$\boldsymbol{x}_4$
$\boldsymbol{x}_5$

Search Space

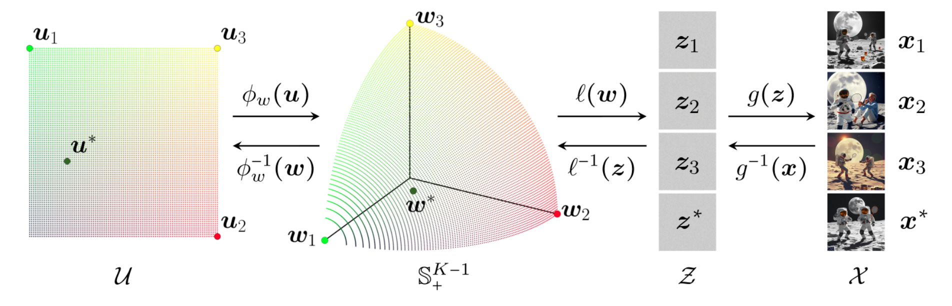
P1 Validity All locations (in the surrogate space) must be supported by the generative model.

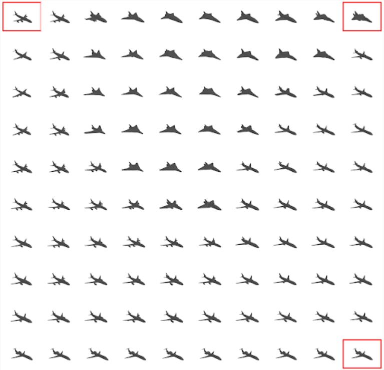
P2 Uniqueness All locations must encode unique objects given the seeds.

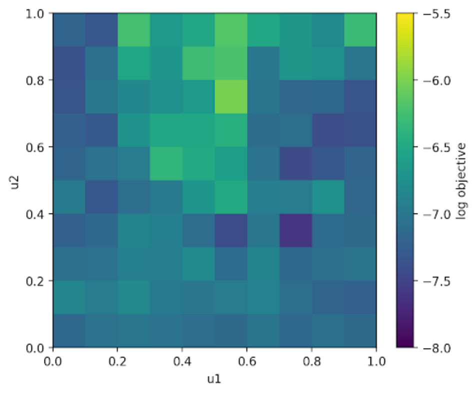
P3 Stationarity The relationship between objects' similarity as a function of their Euclidean distance in the surrogate space should be approximately maintained for any pair of objects throughout the space

Search Space

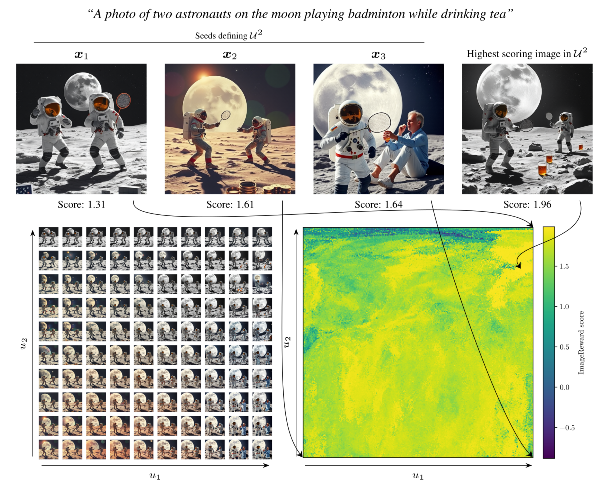
$$\phi : \mathcal{U} \to \mathcal{Z}$$

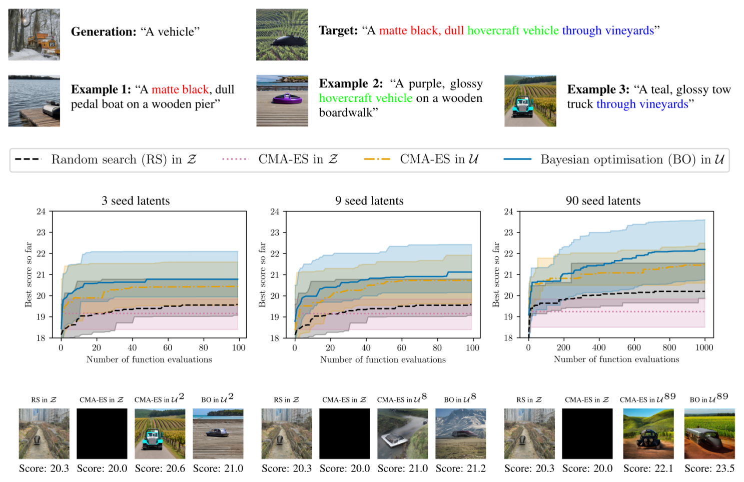

$\boldsymbol{u}_1$
$\boldsymbol{u}_3$
$\boldsymbol{u}^*$
$\boldsymbol{u}_2$
$\mathcal{U}$
$\phi_w(\boldsymbol{u})$
$\phi_w^{-1}(\boldsymbol{w})$
$\boldsymbol{w}_3$
$\boldsymbol{w}^*$
$\boldsymbol{w}_1$
$\boldsymbol{w}_2$
$\mathbb{S}_+^{K-1}$
$\ell(\boldsymbol{w})$
$\ell^{-1}(\boldsymbol{z})$
$\boldsymbol{z}_1$
$\boldsymbol{z}_2$
$\boldsymbol{z}_3$
$\boldsymbol{z}^*$
$\mathcal{Z}$
$g(\boldsymbol{z})$
$g^{-1}(\boldsymbol{x})$
$\boldsymbol{x}_1$
$\boldsymbol{x}_2$
$\boldsymbol{x}_3$
$\boldsymbol{x}^*$
$\mathcal{X}$

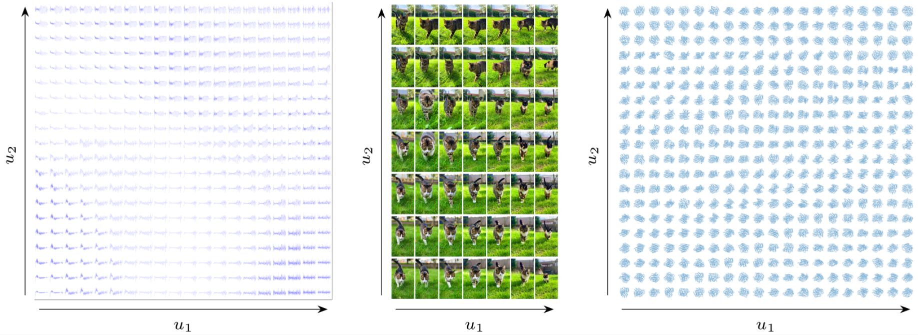


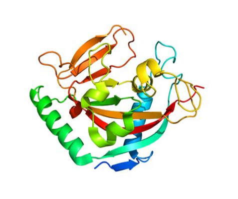


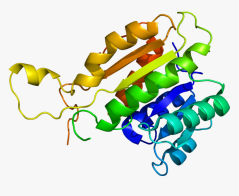
1.0
0.8
0.6
0.4
0.2
0.0
u2
0.0
0.2
0.4
0.6
0.8
1.0
u1
−5.5
−6.0
−6.5
−7.0
−7.5
−8.0
log objective

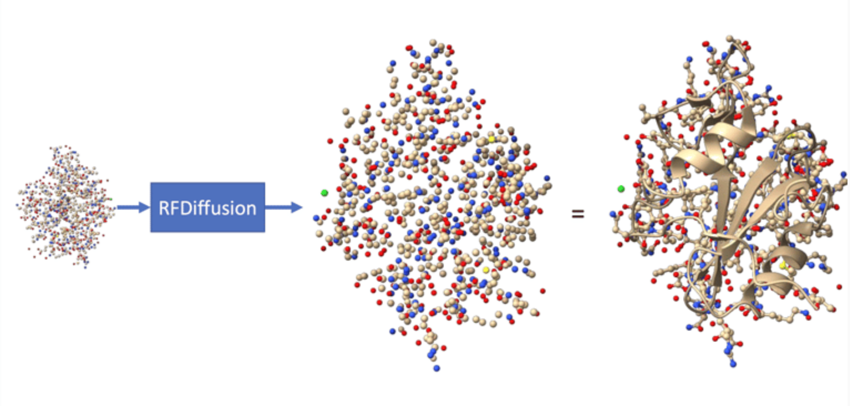

"A photo of two astronauts on the moon playing badminton while drinking tea"
Seeds defining $\mathcal{U}^2$
$\boldsymbol{x}_1$
$\boldsymbol{x}_2$
$\boldsymbol{x}_3$
Highest scoring image in $\mathcal{U}^2$
Score: 1.31
Score: 1.61
Score: 1.64
Score: 1.96
u_2
u_1
u_2
u_1
ImageReward score
1.5
1.0
0.5
0.0
−0.5

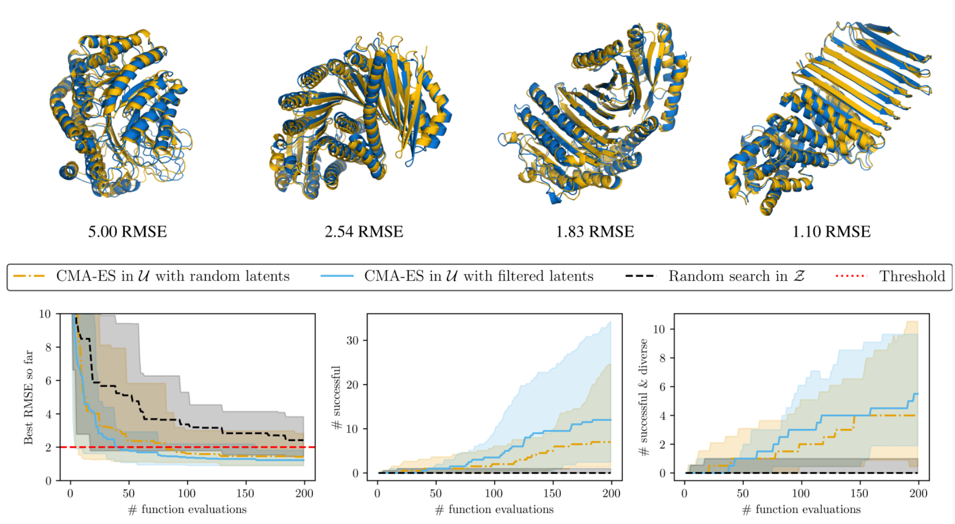

Generation: "A vehicle"
Target: "A matte black, dull hovercraft vehicle through vineyards"
Example 1: "A matte black, dull pedal boat on a wooden pier"
Example 2: "A purple, glossy hovercraft vehicle on a wooden boardwalk"
Example 3: "A teal, glossy tow truck through vineyards"
Random search (RS) in $\mathcal{Z}$
CMA-ES in $\mathcal{Z}$
CMA-ES in $\mathcal{U}$
Bayesian optimisation (BO) in $\mathcal{U}$
3 seed latents
9 seed latents
90 seed latents
Best score so far
Number of function evaluations
RS in $\mathcal{Z}$
CMA-ES in $\mathcal{Z}$
CMA-ES in $\mathcal{U}^2$
BO in $\mathcal{U}^2$
Score: 20.3
Score: 20.0
Score: 20.6
Score: 21.0
RS in $\mathcal{Z}$
CMA-ES in $\mathcal{Z}$
CMA-ES in $\mathcal{U}^8$
BO in $\mathcal{U}^8$
Score: 20.3
Score: 20.0
Score: 21.0
Score: 21.2
RS in $\mathcal{Z}$
CMA-ES in $\mathcal{Z}$
CMA-ES in $\mathcal{U}^{89}$
BO in $\mathcal{U}^{89}$
Score: 20.3
Score: 20.0
Score: 22.1
Score: 23.5


u_2
u_1
u_2
u_1
u_2
u_1

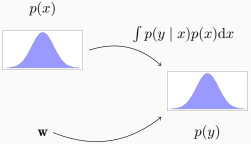





RFDiffusion
=


5.00 RMSE
2.54 RMSE
1.83 RMSE
1.10 RMSE
CMA-ES in $\mathcal{U}$ with random latents
CMA-ES in $\mathcal{U}$ with filtered latents
Random search in $\mathcal{Z}$
Threshold
Best RMSE so far
successful
successful & diverse
function evaluations

Trade-off


$p(x)$
$\int p(y \mid x)p(x)\mathrm{d}x$
$\mathbf{w}$
$p(y)$


$\boldsymbol{x}_1$
$\boldsymbol{x}_2$
$\boldsymbol{x}_3$
$\boldsymbol{x}_4$
$\boldsymbol{x}_5$

Summary

Summary

- a very high-level narrative of recent development

Summary

- a very high-level narrative of recent development
- neural network infrastructure work as black-boxes

Summary

- a very high-level narrative of recent development
- neural network infrastructure work as black-boxes
- constraints/bias comes from algorithms and data

Summary

- a very high-level narrative of recent development
- neural network infrastructure work as black-boxes
- constraints/bias comes from algorithms and data
- Lots of intetersting work in connection with the recent developments and material in this course

eof

References

Bodin, Erik et al. (2024). "Linear Combinations of Gaussian Latents in Generative Models: Interpolation and Beyond". In: *Proceedings of the Thirteenth International Conference on Learning Representations (ICLR)*.

Campbell, NDF and J Kautz (July 2014). "Learning a manifold of fonts". In: *ACM Transactions on Graphics (TOG)* 33.4, p. 91.

Holderrieth, Peter and Ezra Erives (2025). "An Introduction To Flow Matching and Diffusion Models". In: *CoRR*.

Kingma, D. P. and M. Welling (2014). "Auto-Encoding Variational Bayes". In: *Proceedings of the International Conference on Learning Representations*.

Kobyzev, Ivan, Simon J. D. Prince, and Marcus A. Brubaker (2019). "Normalizing Flows: an Introduction and Review of Current Methods". In: *CoRR*.

Lawrence, Neil D. (2004). "Gaussian Process Models for Visualisation of High Dimensional Data". In: *Advances in Neural Information Processing Systems*. Ed. by Sebastian Thrun, Lawrence Saul, and Bernhard Schölkopf. Vol. 16. Cambridge, MA: MIT Press, pp. 329–336.

Meyer, Jean-Arcady and Stewart W. Wilson (1991). "A Possibility for Implementing Curiosity and Boredom in Model-Building Neural Controllers". In: *From Animals to Animats: Proceedings of the First International Conference on Simulation of Adaptive Behavior*, pp. 222–227.

Tipping, Michael E and Christopher M Bishop (1999). "Probabilistic principal component analysis". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 61.3, pp. 611–622.

Willis, Samuel et al. (2025). "Define Latent Spaces By Example: Optimisation Over the Outputs of Generative Models". In: *CoRR*.