Artem Ryzhikov

Autoencoders

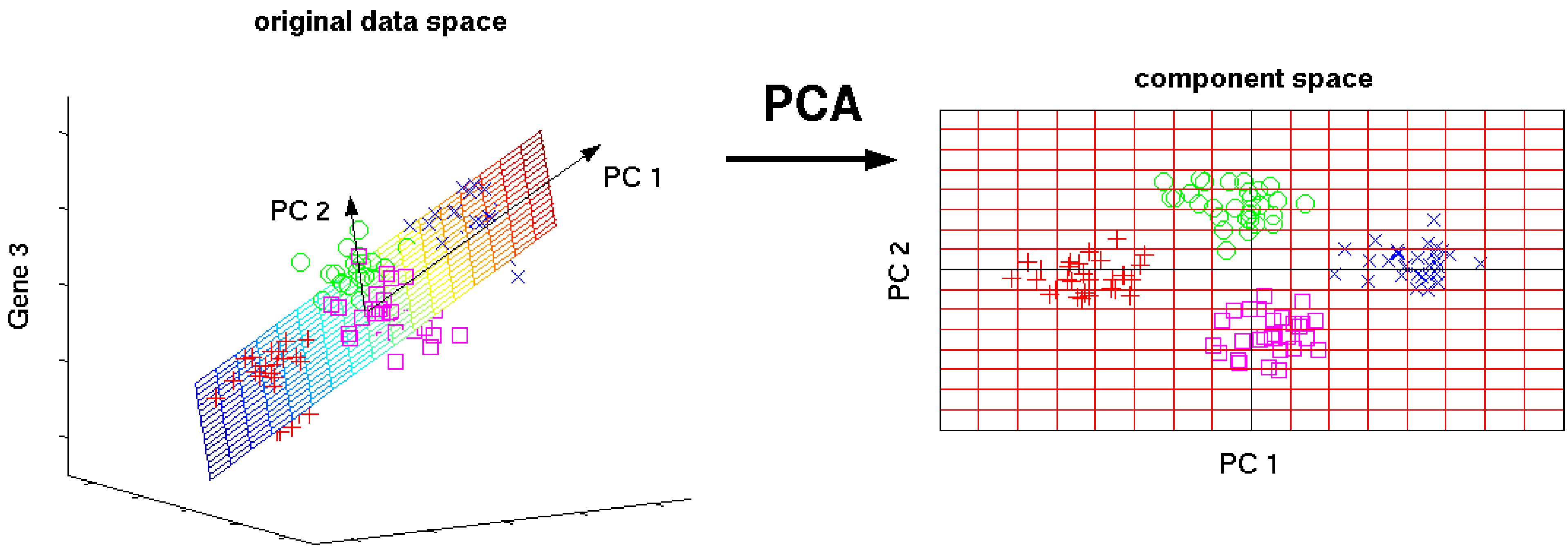
2021

MLHEP 2021
Higher School of Economics
National Research University
LAMBDA • HSE
Yandex
School of Data Analysis
EPFL
Schaffhausen Institute of Technology

Principal component analysis (PCA)

Principal component analysis (PCA)

Goal: fit transformation $X' = AX$, where $X \in \mathbb{R}^n, X' \in \mathbb{R}^k$, n is original dimension, $k \leq n$ is a number of principal components, $A \in \mathbb{R}^{k \times n}$ is linear transformation matrix to basis of principal components



PCA. Pros and cons.

Advantages:

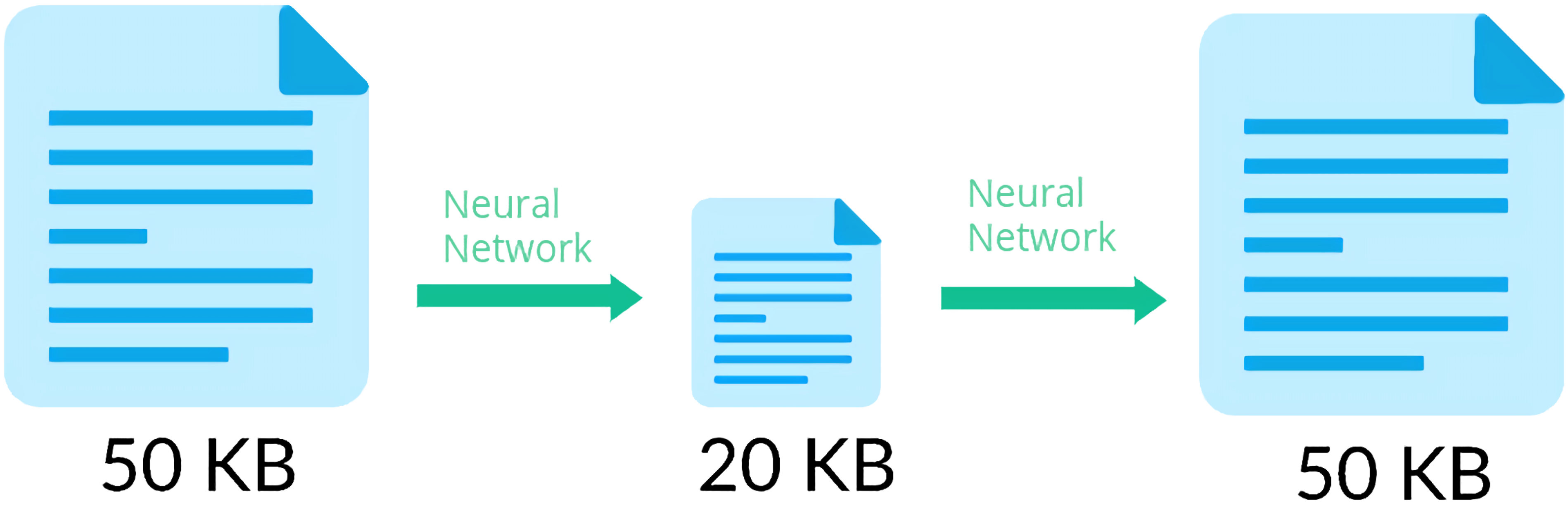
- **Optimal low-rank approximation** in terms of squared loss
- New features (principal components) are **uncorrelated**
- **Importance** (eigenvalues) of that new features is automatically obtained
- Only the most principal ones can be taken to **reduce dimensionality without** significant **losses**

Disadvantages:

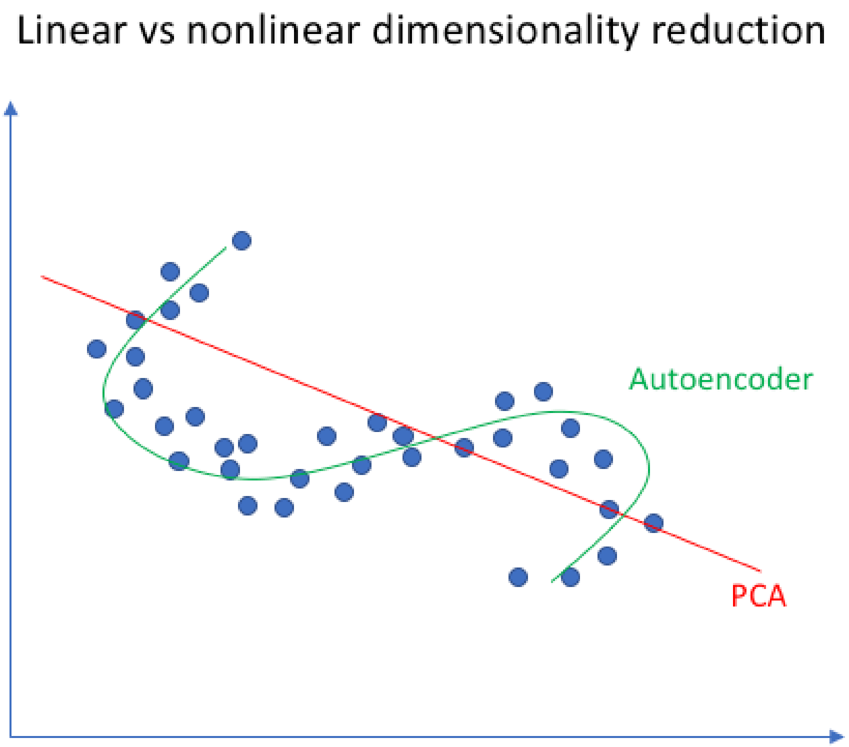
- PCA corresponds to **linear transformation only**
- **Computationally expensive and non-scalable**. Time complexity is $O(nm^2)$ for matrix $X \in \mathbb{R}^{m \times n}$ where $m \leq n$

Autoencoder

Autoencoder. Idea



Autoencoder vs. PCA



PCA:

$X' = AX | A \in \mathbb{R}^{k \times n}, X \in \mathbb{R}^n, X' \in \mathbb{R}^k$

Autoencoder:

$X' = \text{Encoder}(X|\theta) | X \in \mathbb{R}^n, X' \in \mathbb{R}^k$

$\hat{X} = \text{Decoder}(X'|\phi) | \hat{X} \in \mathbb{R}^n, X' \in \mathbb{R}^k$

$\hat{X}$ - reconstructed object, θ and ϕ are Encoder and Decoder parameters respectively.

Autoencoder (AE)

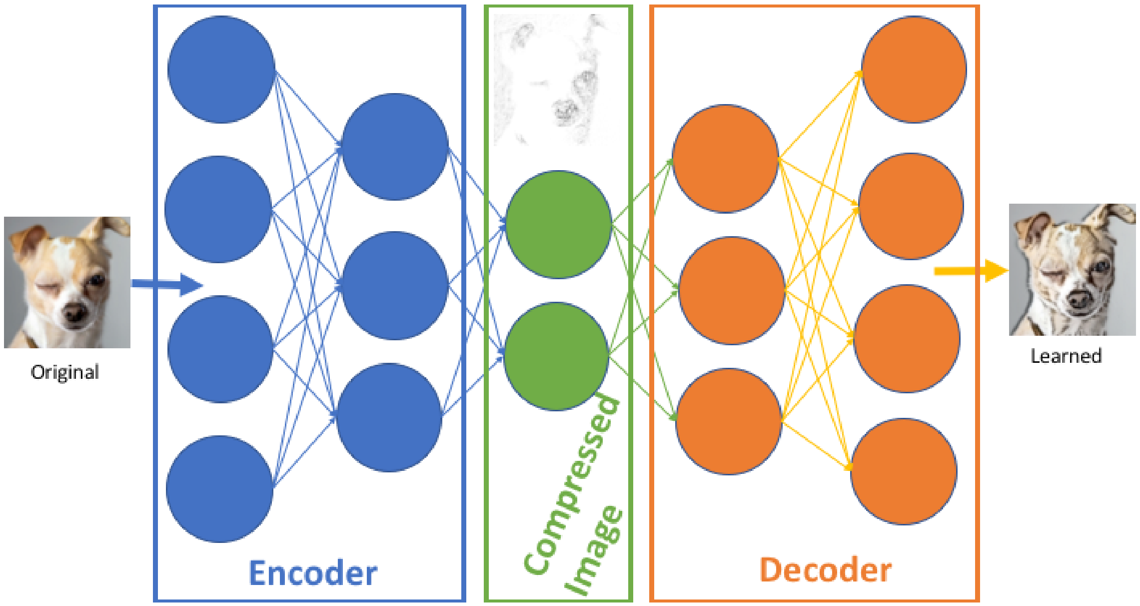
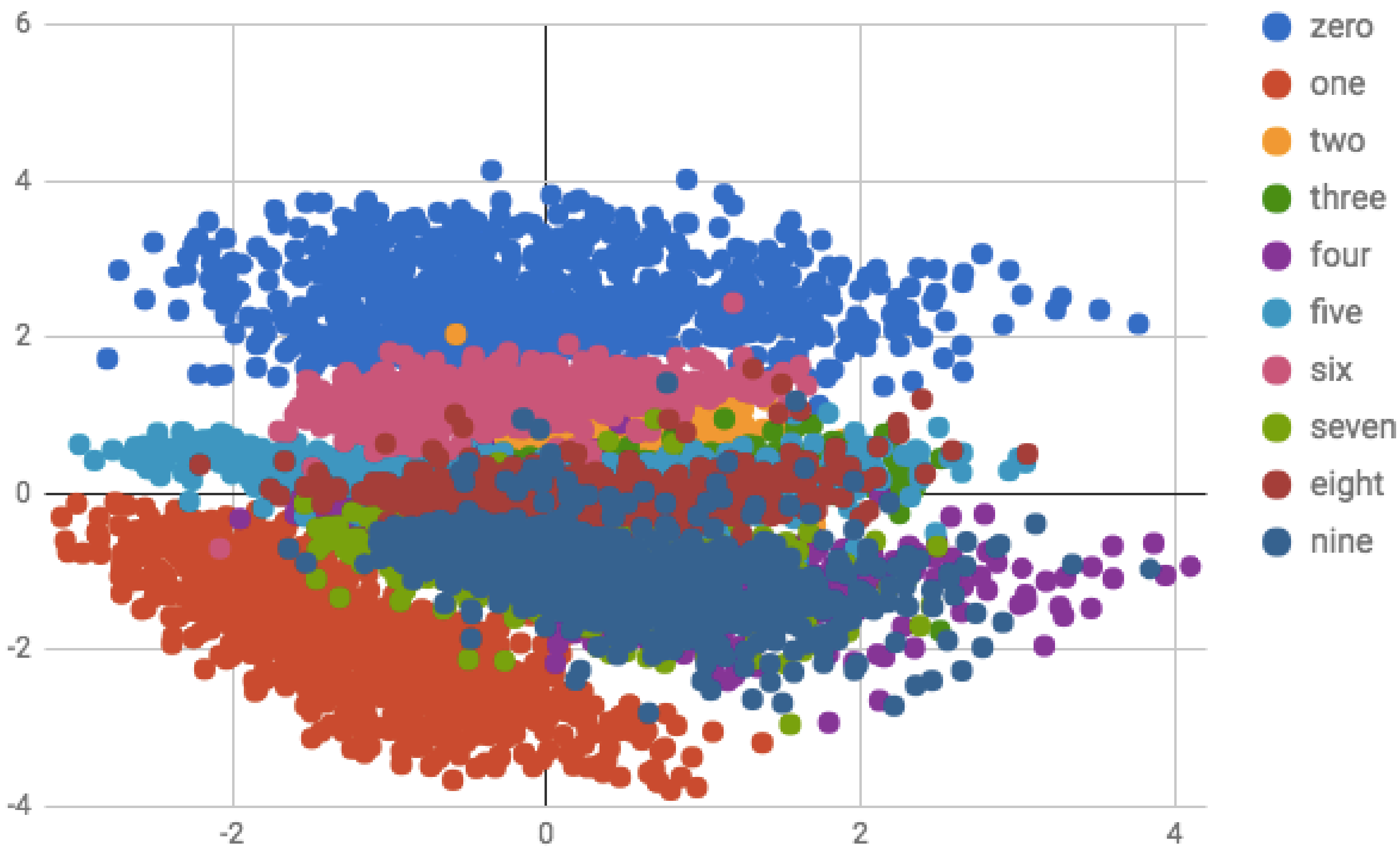


Figure: https://arnoldkokoroko.com/projects/imagecompress/

Train: minimize reconstruction loss between original and reconstructed objects: $\min_{\theta,\phi} \mathcal{L}(\mathsf{X}, \hat{\mathsf{X}})$, where X is original object, $\hat{\mathsf{X}}$ is reconstructed ("uncompressed") object, $\mathcal{L}(\cdot,\cdot)$ is reconstruction loss, θ and ϕ are Encoder and Decoder parameters respectively

Autoencoder. Example (MNIST)



Autoencoder. Physics example

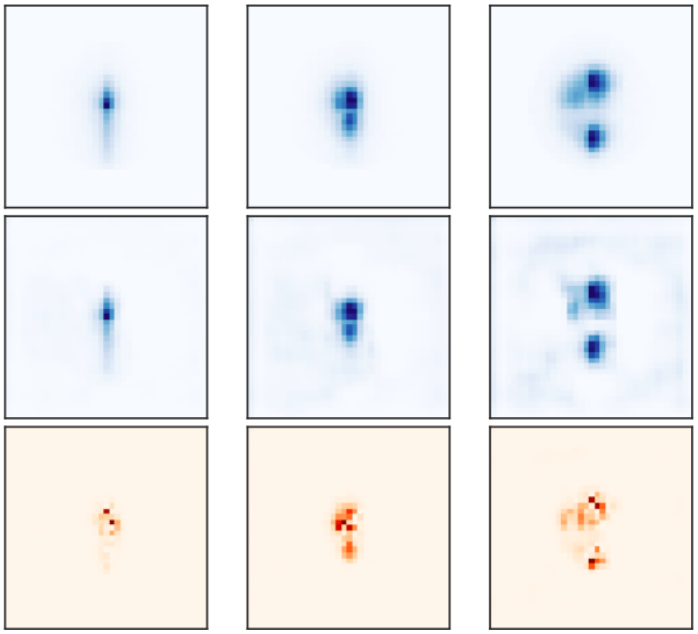


Figure 3: Each panel represents the average of 100k jet images. Pixel intensity corresponds to the total p_T in each pixel. Upper row: original sample. Middle row: after reconstruction. Lower row: pixel-wise squared error. Left column: QCD jets. Middle column: top jets. Right column: $\tilde{g}$ jets.

M. Farina et al., "Searching for New Physics with Deep Autoencoders"

Sparse autoencoders

Sparse autoencoder

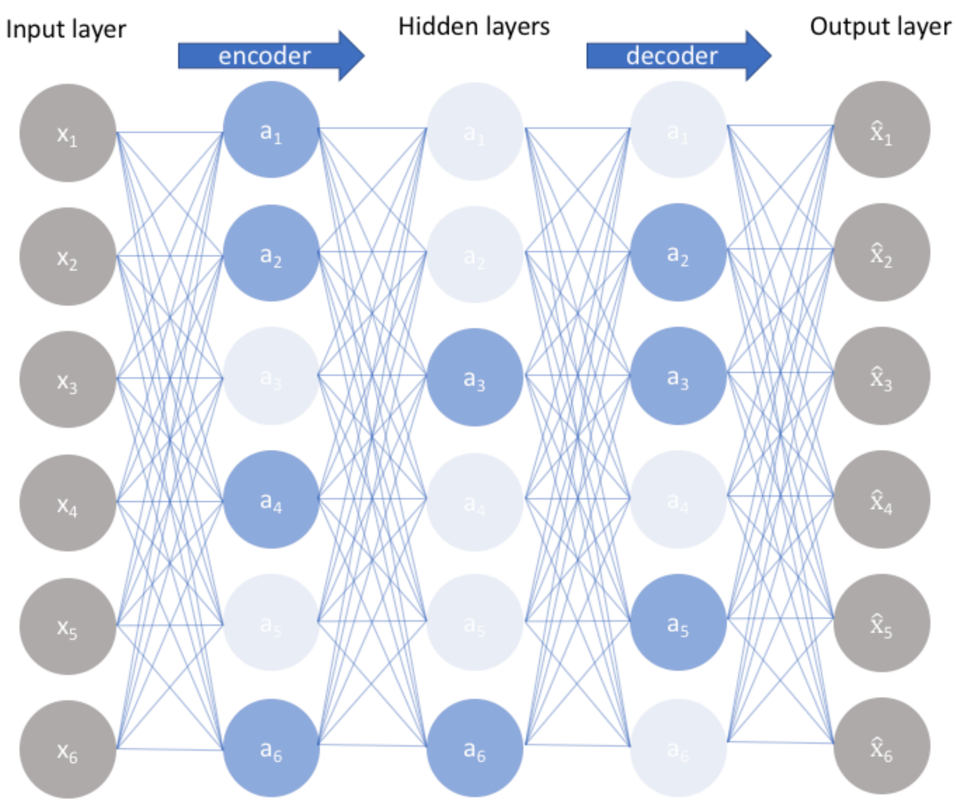


Figure: https://www.jeremyjordan.me/autoencoders/

Problem Usually compressed data from autoencoder is still redundant. We need to force autoencoder to use sparsified hidden representation of data

Sparse autoencoder. Regularization

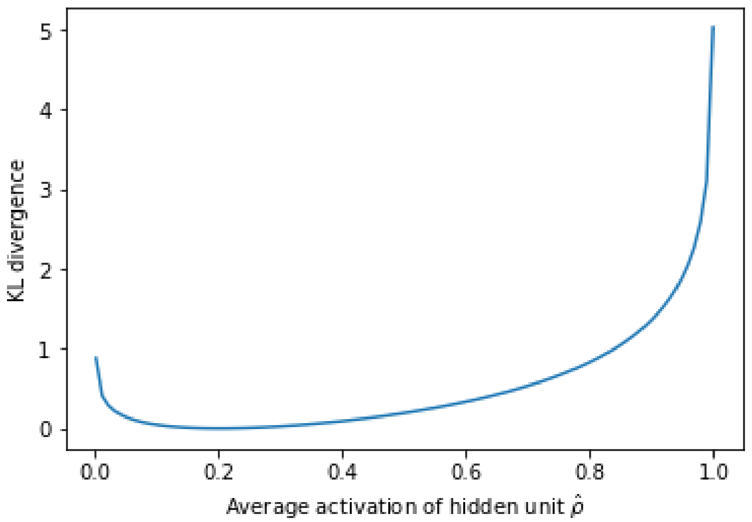
- Autoencoder: $\min_{\theta,\phi} \mathcal{L}(X, \hat{X})$
- Sparse autoencoder:

$\min_{\theta,\phi}[\mathcal{L}(X, \hat{X}) + \text{regularization}(\theta)]$

Two kinds of regularization in Sparse Autoencoder:

- L1-regularization: $\mathcal{L}\left(X, \hat{X}\right) + \lambda \sum_i \left|a_i^{(h)}\right|$
- KL divergence with Bernoulli distribution:

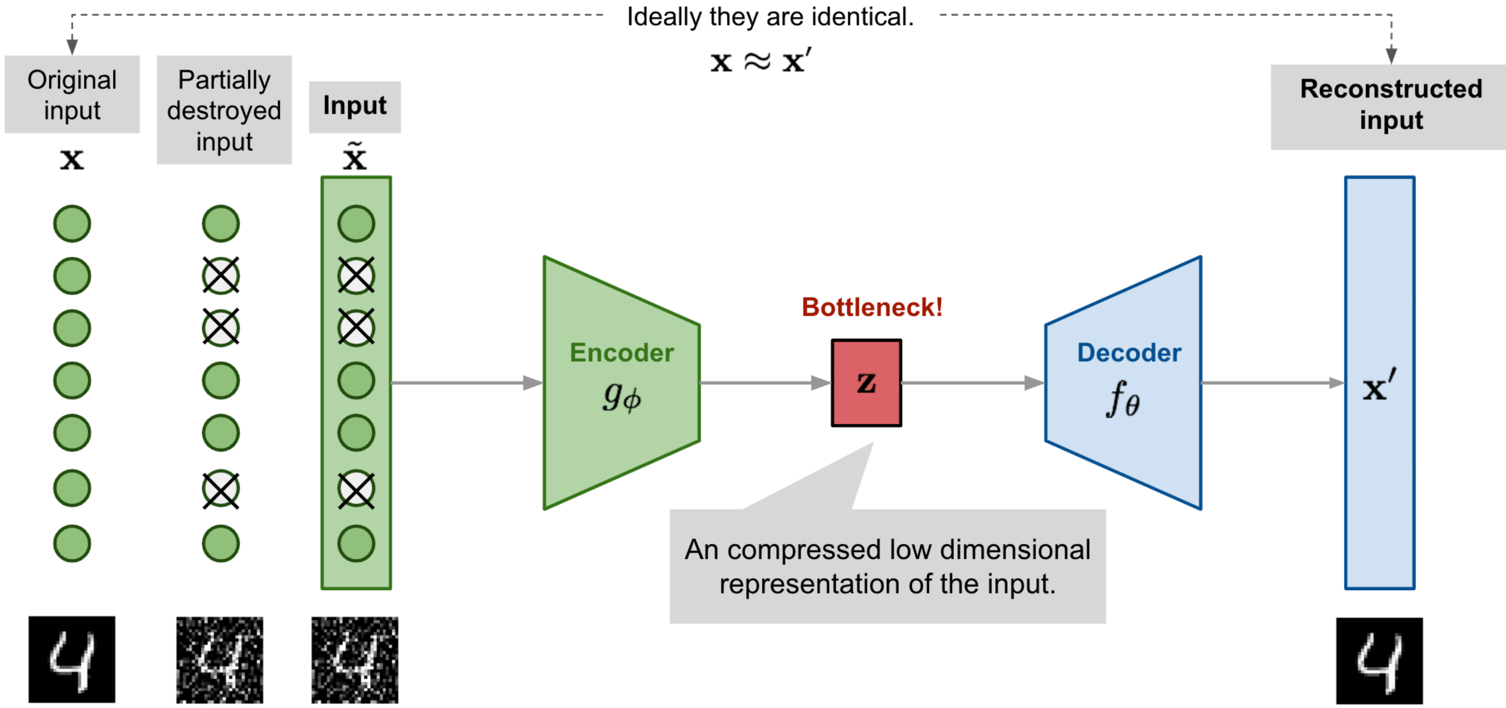
$\mathcal{L}\left(X, \hat{X}\right) + \sum_j KL\left(\rho || \hat{\rho}_j\right)$



Denoising autoencoder

Denoising autoencoder (DAE)

Problem Usually compressed data from autoencoder is still redundant (which leads to overfitting). We need to force AE to use sparsified hidden representation of data



Idea: corrupt input to prevent autoencoder from overfitting

Variational Autoencoder

Variational Autoencoder (recap). Motivation

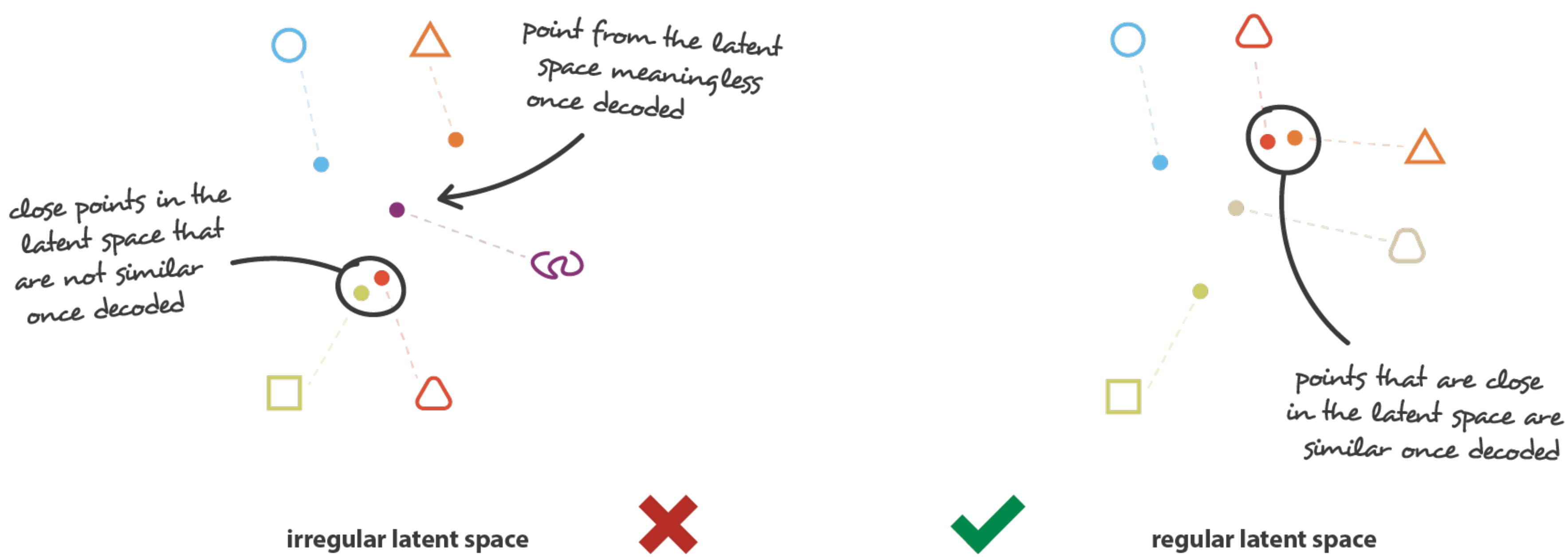


Figure: https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73

Variational Autoencoder (recap). Idea

- Autoencoder: $\hat{\mathsf{X}} = \mathsf{Decoder}(\mathsf{Encoder}(\mathsf{X}|\theta)|\phi)$
- Variational autoencoder: $\hat{\mathsf{X}} = \mathsf{Decoder}(\mathsf{Z}|\phi)$, where $\mathsf{Z} \sim \mathsf{q}(\mathsf{Z}|\mathsf{X}, \theta) = \mathsf{Encoder}(\mathsf{X}|\theta)$

Problem: how to fit θ and ϕ?

$$\log[\mathsf{Pr}(\mathbf{X}|\boldsymbol{\phi})] = \log[\int \mathsf{Pr}(\mathbf{X}, \mathbf{Z}|\boldsymbol{\phi})\mathsf{d}\mathbf{Z}] = \log[\int \mathsf{q}(\mathbf{Z}|\mathsf{X}, \theta)\frac{\mathsf{Pr}(\mathbf{X}, \mathbf{Z}|\boldsymbol{\phi})}{\mathsf{q}(\mathbf{Z}|\mathbf{X}, \theta)}\mathsf{d}\mathbf{Z}] \geq$$

$$\geq \int \mathsf{q}(\mathbf{Z}|\mathsf{X}, \theta)\log[\frac{\mathsf{Pr}(\mathbf{X}, \mathbf{Z}|\boldsymbol{\phi})}{\mathsf{q}(\mathbf{Z}|\mathbf{X}, \theta)}]\mathsf{d}\mathbf{Z} = \mathsf{ELBO}(\theta, \phi)$$

$$\max_{\phi} \log[\mathsf{Pr}(\mathbf{X}|\boldsymbol{\phi})] \rightarrow \max_{\theta, \phi} \mathsf{ELBO}(\theta, \phi)$$

ELBO

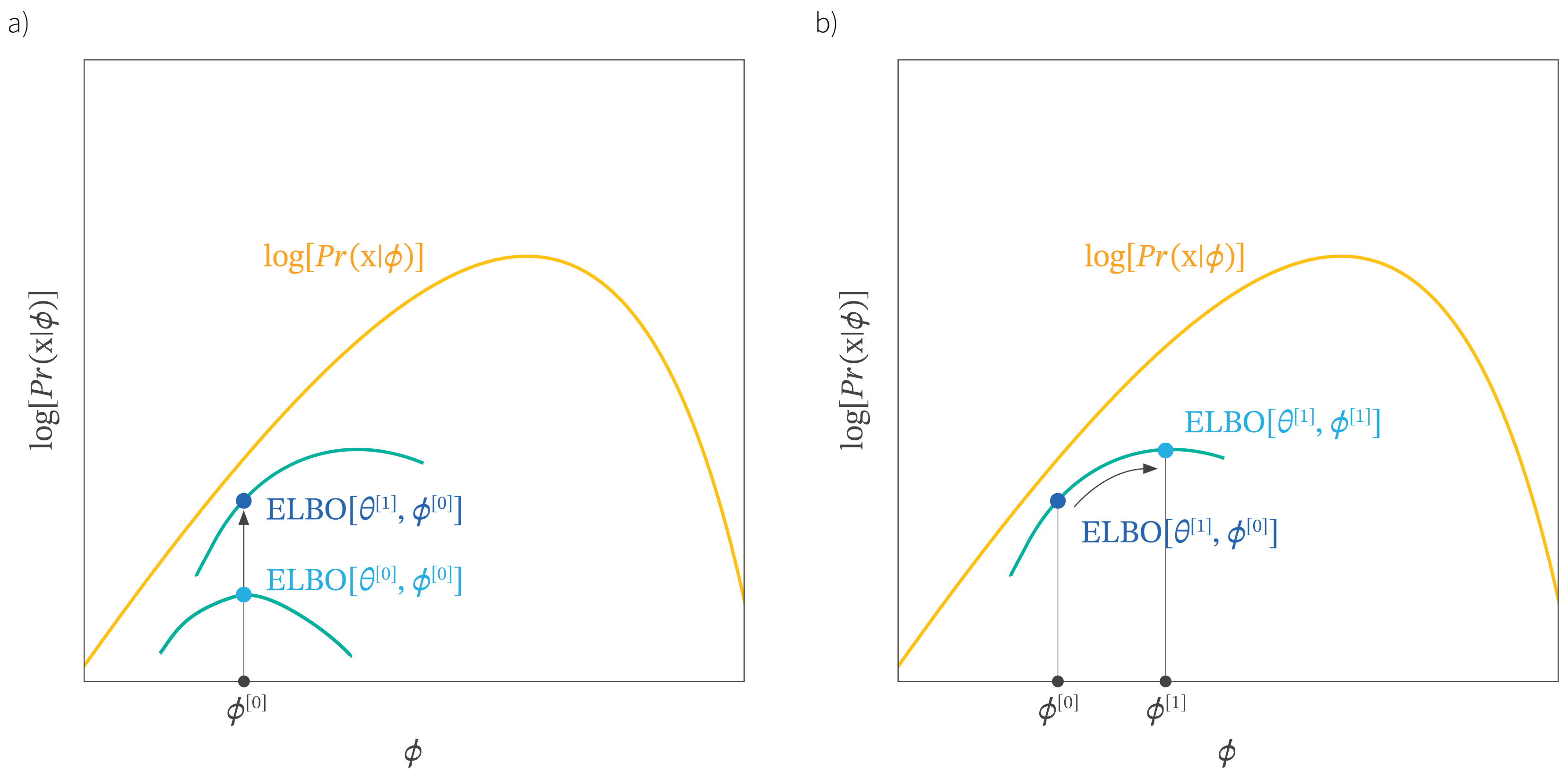


Figure: https://www.borealisai.com/en/blog/tutorial-5-variational-auto-encoders/

ELBO

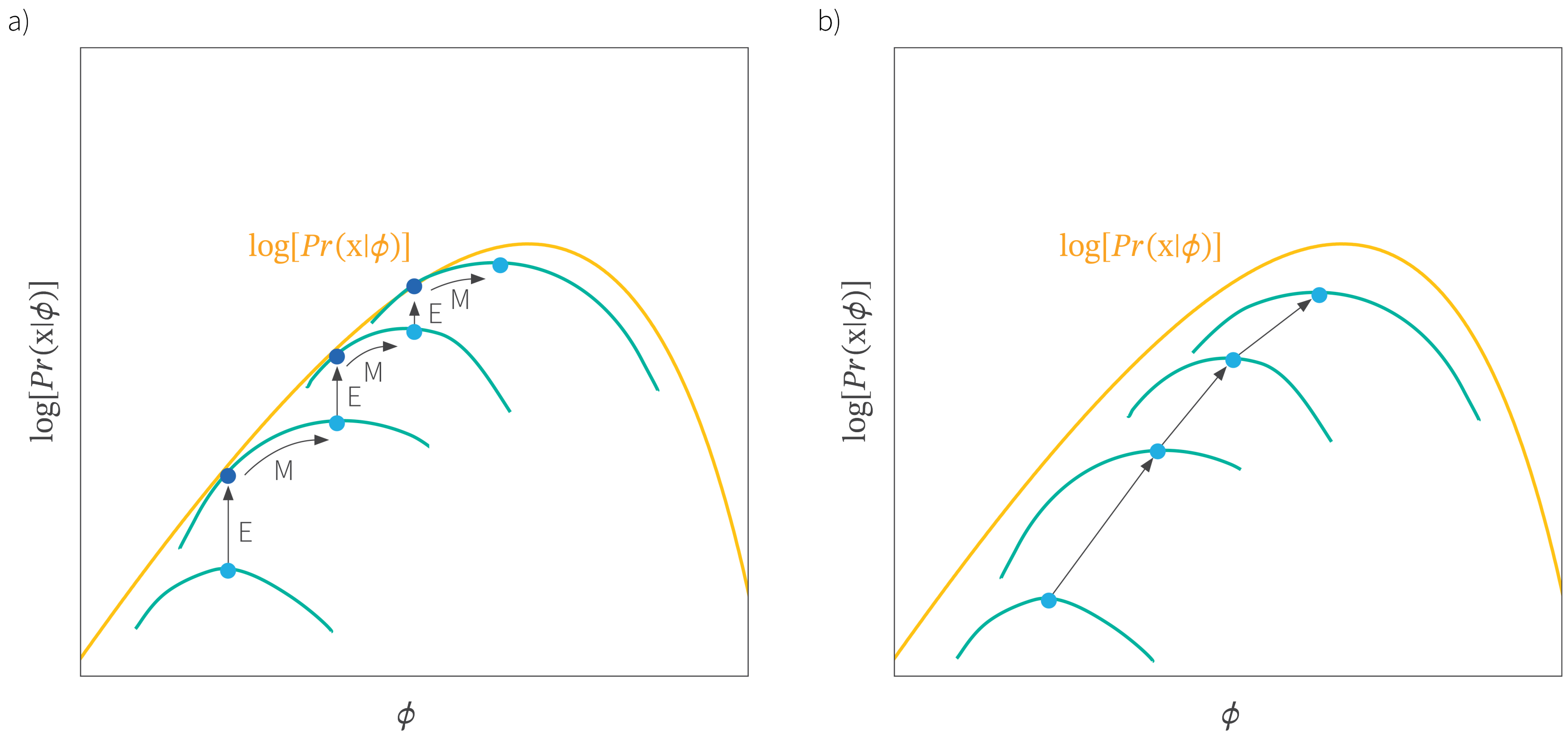


Figure: https://www.borealisai.com/en/blog/tutorial-5-variational-auto-encoders/

Conditional VAE

CVAE

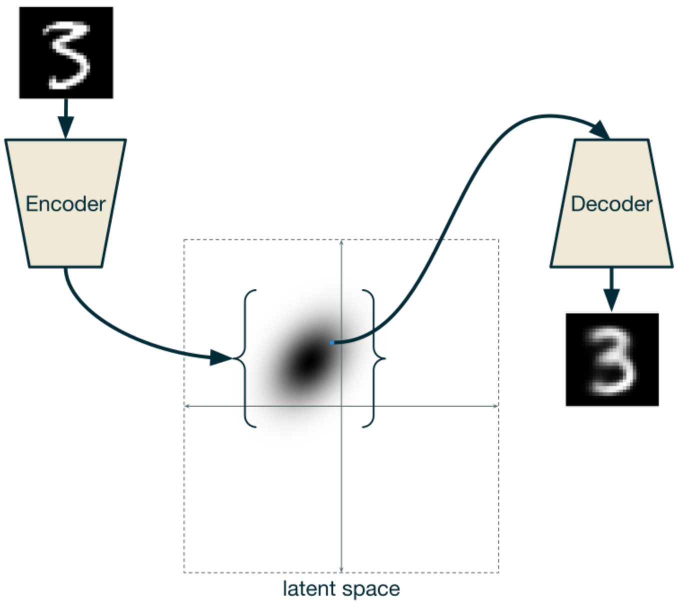


Figure 1: VAE. Label is not used

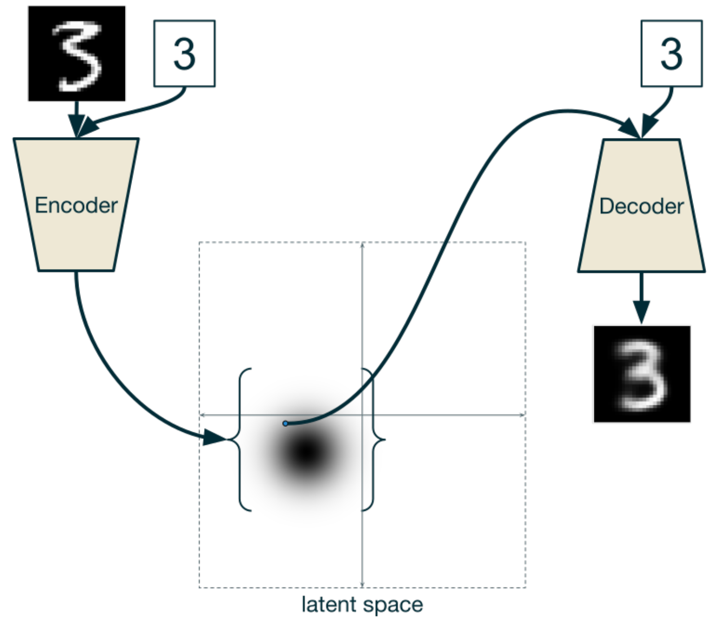


Figure 2: Conditional VAE. Label as extra condition is used

Thank you for your attention!

Artem Ryzhikov aryzhikov@hse.ru