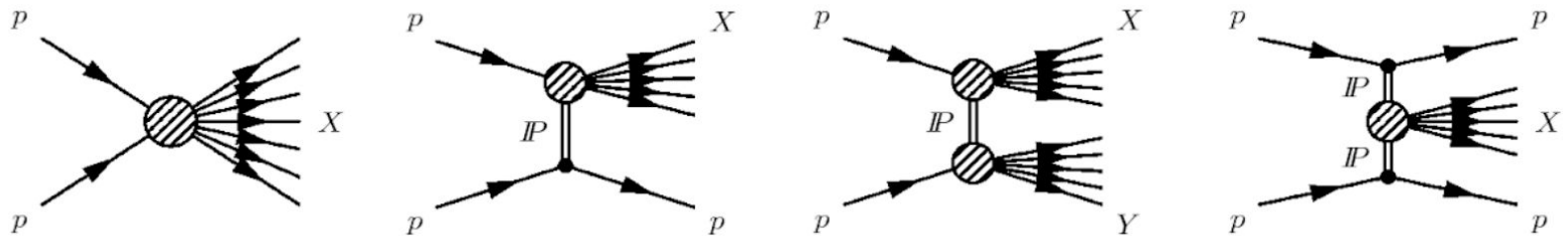
Unsupervised classification of particle interaction events using deep neural network

Rafał Staszewski, Marcin Wolter

4 July 2022

Task

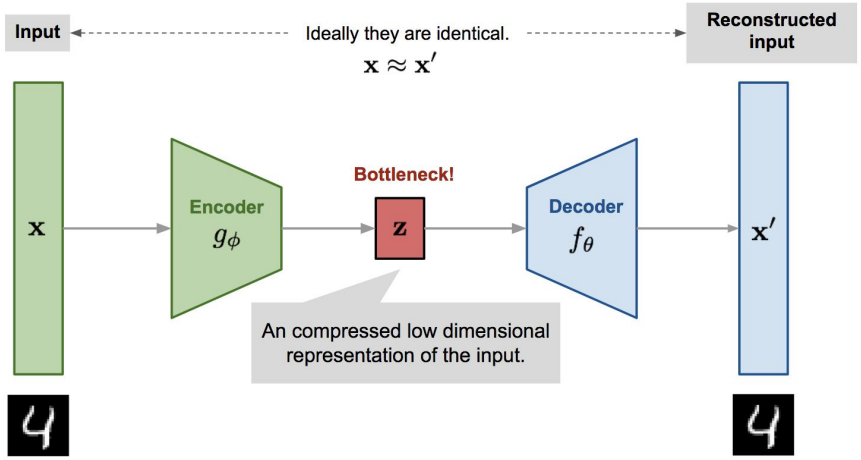
- The task is to perform unsupervised classification of simulated physics processes:



- The idea is to do it in steps:
  - Reduction of dimensionality using deep neural network as an autoencoder,
  - Finding clusters in the reduced space using one of the clustering algorithms (for example kMeans)
- Before performing the analysis on physics data we should test the entire chain on some data, which are easier to see for human beings, for example MNIST dataset of hand-written digits.

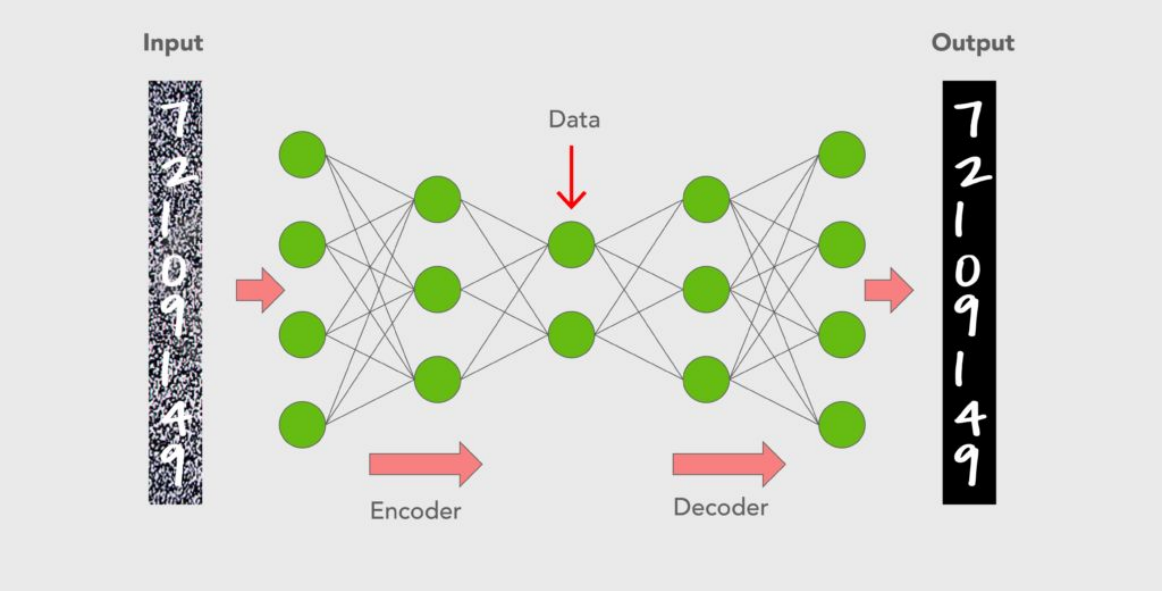
Autoencoder

- Autoencoders are used to reduce the data dimensionality while preserving the maximum of information.



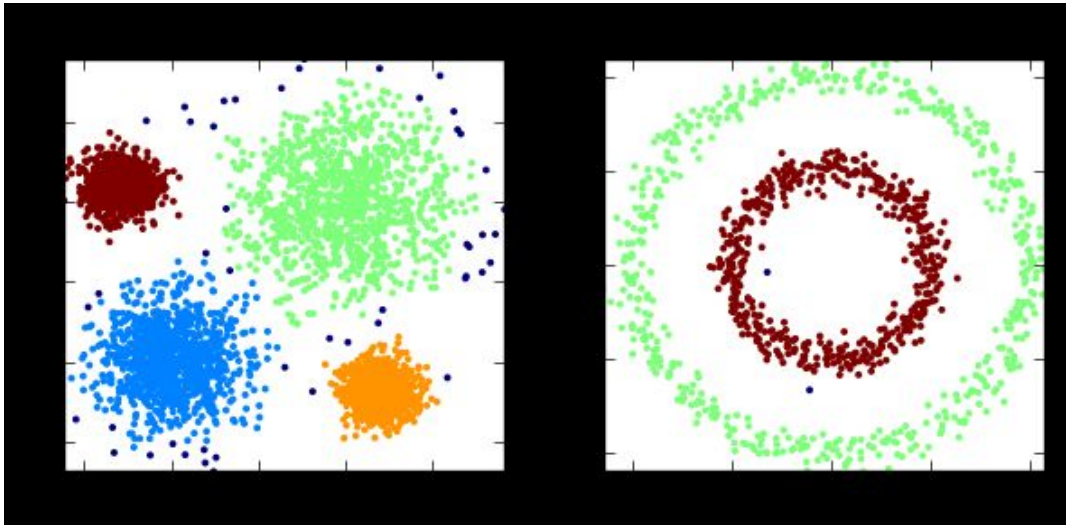
- The network is trained to reproduce input on output. But information must go through the bottleneck.
- Result: encoder encodes input into low dimensional space, decoder decodes it again.

Autoencoder



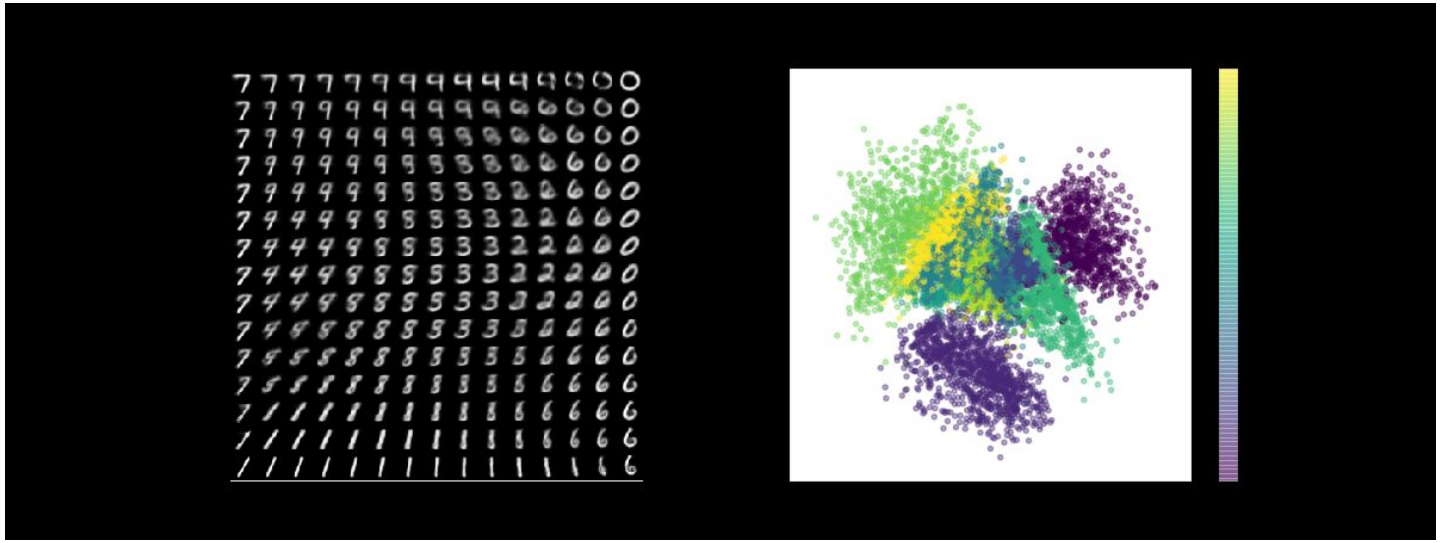
- Autoencoders might be used for noise reduction

Clustering



- **Clustering** - task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar (in some sense) to each other than to those in other groups (clusters). It is a main task of exploratory data analysis.
- Clustering in low-dimensional space might allow to group similar events (handwritten digits) into cluster of similar objects.

Variational autoencoder VAE



- VAE is an autoencoder whose encodings distribution is regularised during the training in order to ensure that its latent space has good properties allowing us to generate some new data.
- From any point in 2D latent space we get reasonable letter and the map is pretty regular.
- We might learn how to generate the new events, similar to the original ones.
- **We might use VAE to learn how to simulate physics events and see, which classes are similar to each other (like digits in the above plot).**

Tools

- Tools to be used:
  - Google Colab https://colab.research.google.com – to run the code and collaborate
  - https://github.com - to store the code.
  - Python programming language
  - https://keras.io package and tensorflow to build deep neural networks.
  - https://scikit-learn.org/ - machine learning in python.

Summary

- Unsupervised classification of particle interaction events using deep neural network

- Machine learning is often used in particle physics for the classification of events. The goal of the project is to attempt an unsupervised classification of events, where no information about the type of event is used. The proposed approach is based on an Deep Neural Network (DNN) auto-encoder for the dimensionality reduction together with a clustering algorithm. The procedure starts with Monte Carlo event generators, which are used to produce samples of artificial, pseudo-random events, which are later used for DNN training.

- The project will be performed using the Google Colab environment with Tensorflow, Keras and scikit-learn libraries. A good understanding of Python programming is needed.