Unsupervised Learning with Autoencoders and Generative Adversarial Networks

Jeffrey Kelling

27th September 2018

HZDR
HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF

1 Autoencoders

2 Generative Models

- Variational Autoencoders
- Generative Adversarial Networks (GANs)

Member of the Helmholtz Association
Jeffrey Kelling | Department of Information Services and Computing | http//www.hzdr.de

Autoencoders

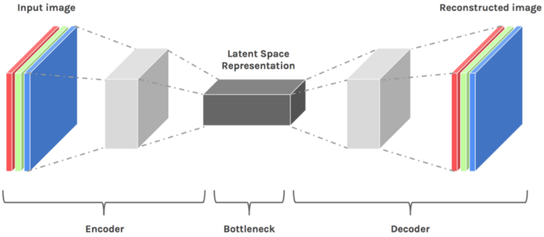
Unspervised Learning. Autoencoders

- Training using unlabelled data

figure: Julien Despois @ medium.com

Unspervised Learning. Autoencoders

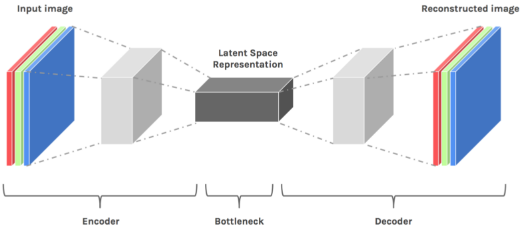
- Training using unlabelled data



- Optimization goal is to reconstruct input image as output
- Bottleneck forces network to learn feature-based representation

figure: Julien Despois @ medium.com

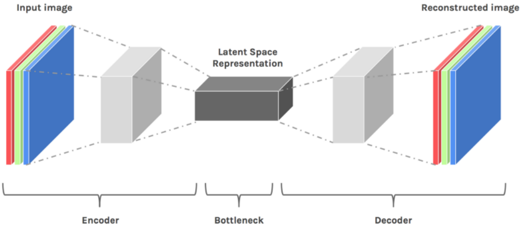
Why?



1. Latent space smaller tahn input ⇒ compression
   - errors hard to control

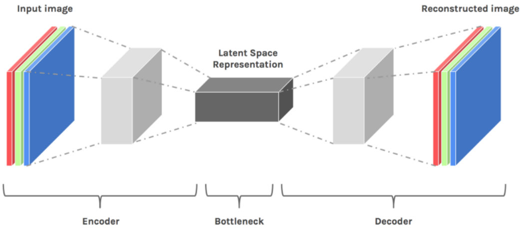
Member of the Helmholtz Association
Jeffrey Kelling | Department of Information Services and Computing | http//www.hzdr.de

Why?



1. Latent space smaller tahn input ⇒ compression
   - errors hard to control
2. Discovery of frequent patterns in data
   - what gets a place in latent space is common
   - anomaly-detection: rare samples will have high reconstruction errors

Why?



1. Latent space smaller tahn input $\Rightarrow$ compression
   – errors hard to control
2. Discovery of frequent patterns in data
   - what gets a place in latent space is common
   - anomaly-detection: rare samples will have high reconstruction errors
3. Discovery of features with convolutional autoencoders
   - Use encoder as pretrained part of classification of other network

Member of the Helmholtz Association
Jeffrey Kelling | Department of Information Services and Computing | http//www.hzdr.de

Unspervised Learning—Google Brain I

- Deep convolutional autoencoder trained using images from "the internet"[1]

[1] Le, Ranzato et al. 2011

Unspervised Learning—Google Brain I

- Deep convolutional autoencoder trained using images from "the internet"[1]

One neuron in the bottleneck
reacts strongly to faces ...

[1] Le, Ranzato et al. 2011

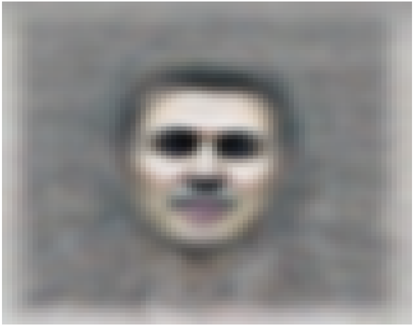
Jeffrey Kelling | Department of Information Services and Computing | http//www.hzdr.de

Unspervised Learning—Google Brain I

- Deep convolutional autoencoder trained using images from "the internet"[1]

One neuron in the bottleneck reacts strongly to faces ...

... it is most strongly excited by this face:



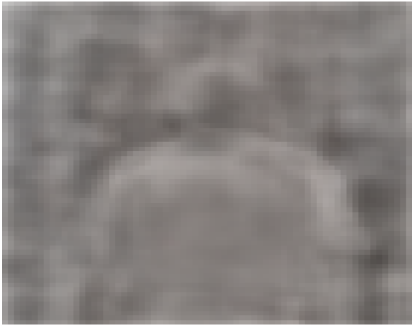
[1] Le, Ranzato et al. 2011

Unspervised Learning—Google Brain II

- Concepts common in the training data automatically learned

cat face

human body



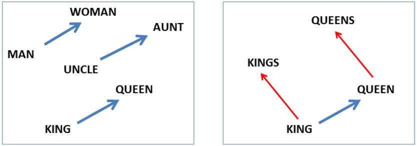
Exercise 1: Autoencoder

`day4/notebooks/MNISTAutoencoder`

Member of the Helmholtz Association
Jeffrey Kelling | Department of Information Services and Computing | http//www.hzdr.de

Specialized embedding algorithms

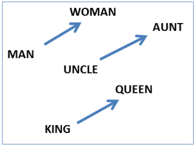
- GloVe https://nlp.stanford.edu/projects/glove/
- word2vec https://arxiv.org/abs/1301.3781

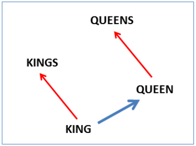


https://www.aclweb.org/anthology/N13-1090

Specialized embedding algorithms

- GloVe https://nlp.stanford.edu/projects/glove/
- word2vec https://arxiv.org/abs/1301.3781





https://www.aclweb.org/anthology/N13-1090

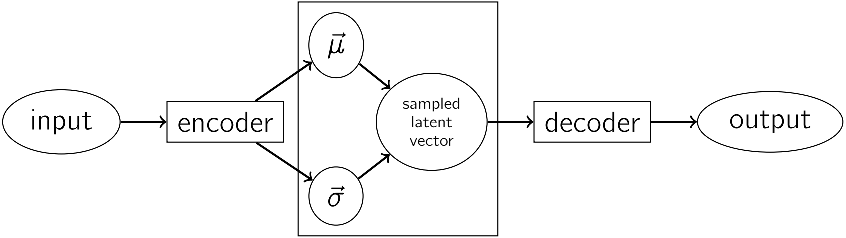
- Uniform Manifold Approximation and Projection (umap) https://github.com/lmcinnes/umap

Generative Models

Member of the Helmholtz Association
Jeffrey Kelling | Department of Information Services and Computing | http//www.hzdr.de

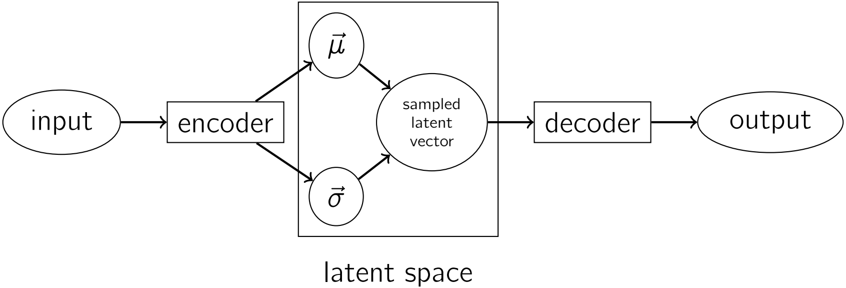
Variational Autoencoders I

- autoencoder which learns the distribution of (input) latent space samples
  - assuming multi-dimensional gaussian
  - learning vectors mean $\vec{\mu}$ and standard deviation $\vec{\sigma}$
- learned distribution is sampled to generate output
  $\Rightarrow$ generative model



latent space

Variational Autoencoders I



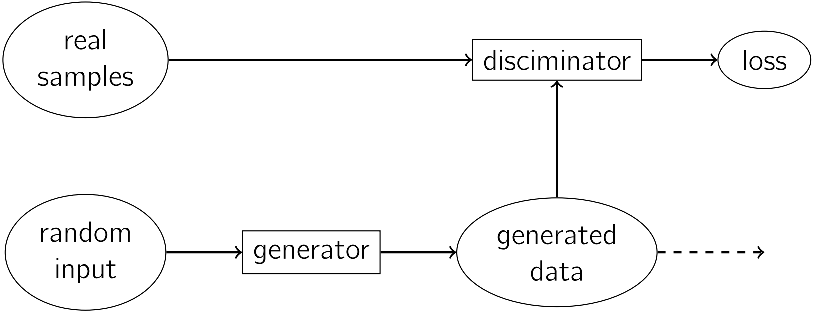
- loss needs to maximize reconstruction and gaussianity of input latent space vectors

$$\text{loss} = \text{reconLoss} + \sum \text{KLDivergence}(\mu_i, \sigma_i)$$

Member of the Helmholtz Association
Jeffrey Kelling | Department of Information Services and Computing | http//www.hzdr.de

Generative Adversarial Networks (GANs)

- two networks competing in a zero-sum game during training
  - D Discriminator: distingiush between **real** and **generated** input
  - G Generator: generate samples, which the discriminator labels as **real**



- also as modified loss function, e.g. when training auto-encoders

Exercise 2: Variational Autoencoder

`day4/notebooks/MNISTVAE`

Member of the Helmholtz Association
Jeffrey Kelling | Department of Information Services and Computing | http//www.hzdr.de