Machine Learning

Prof. Sergei Gleyzer

Lecture

PH451, PH551

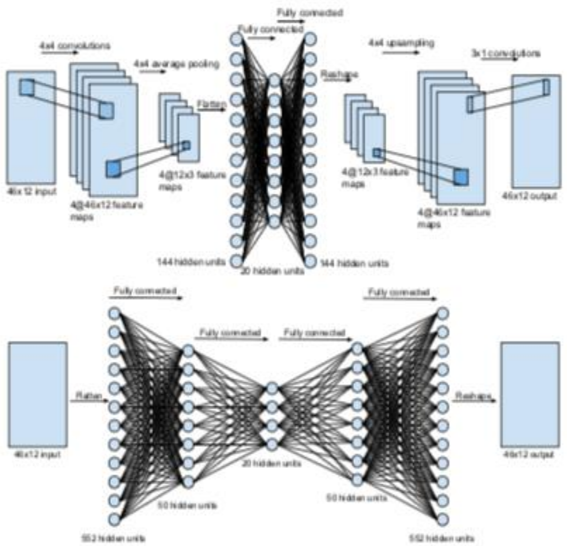
March 27, 2025

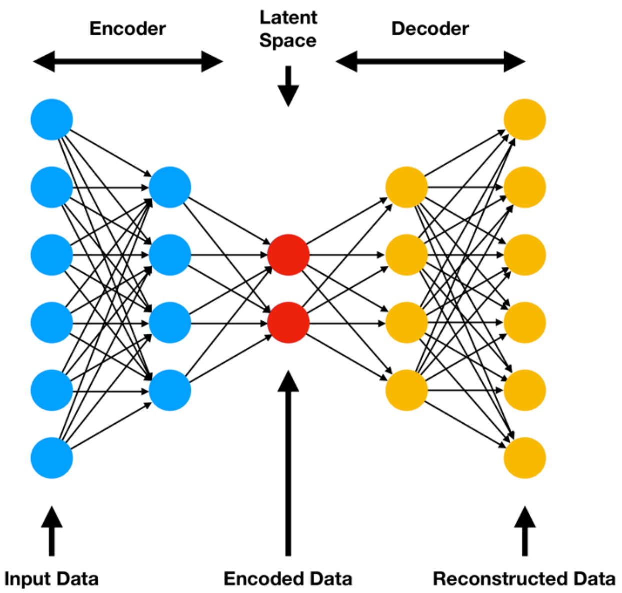
Representation Learning

Auto-Encoders

Learn self-representation

- Rumelhalt et al. (1987, possibly earlier)
- Neural network architecture with a bottleneck structure

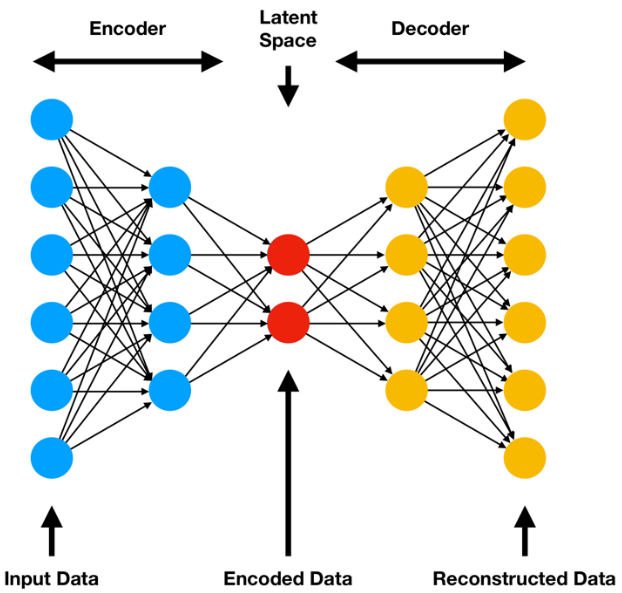




Auto-Encoders

Goal:

- Learn the latent representation (unsupervised)
- Dimensionality reduction (low-dim)
- Feature detection
- Anomaly detection
- Generative models



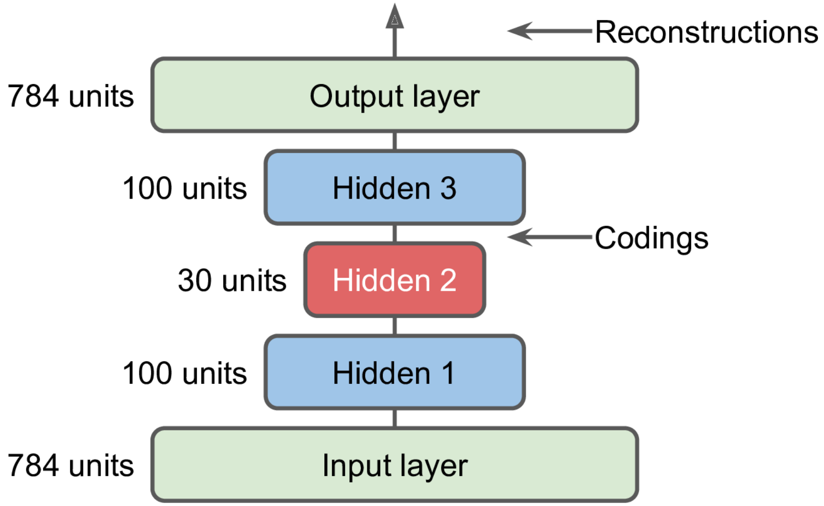
PCA: Linear Auto-Encoder

PCA:

- Linear activation
- MSE Loss

AutoEncoders

- Generally non-linear
- Stacked (add layers)
- Binary Cross-Enthropy Loss



Auto-Encoders

Convolutional AE

- Images

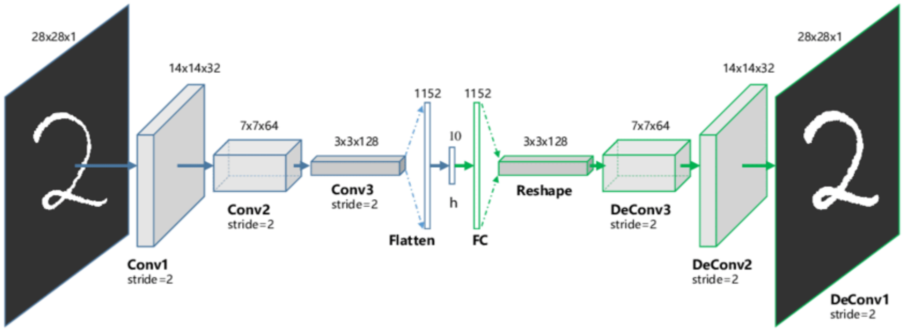
Recurrent AE

- Sequences

Denoising AE

- Add noise, try to recover original

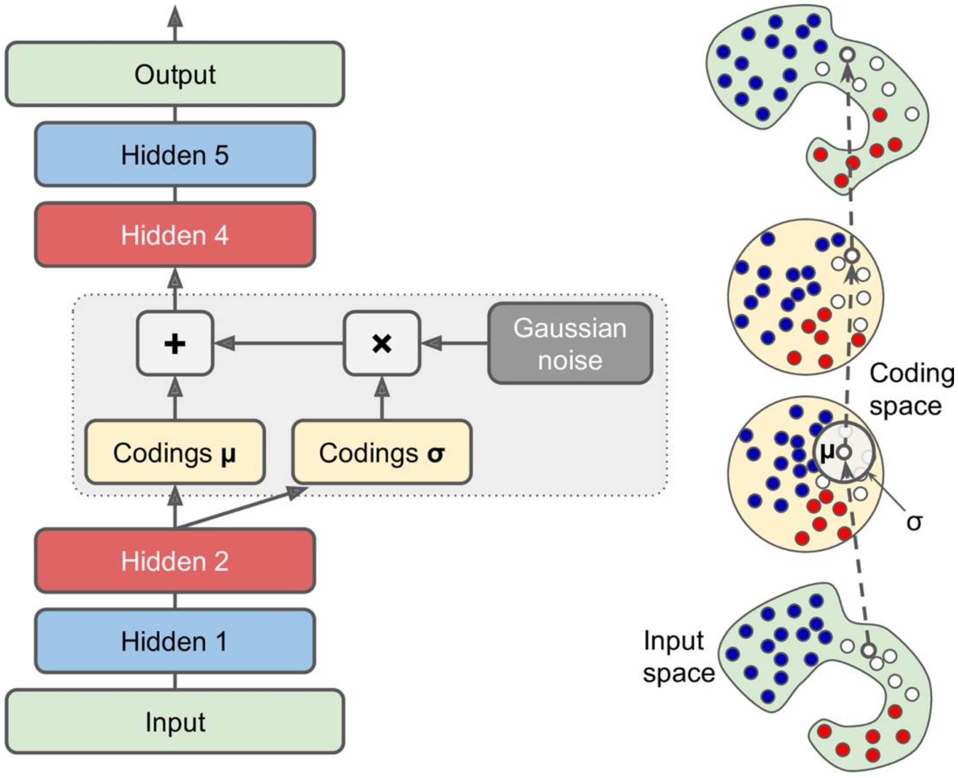
ConvAE



Variational Auto-Encoders

VAE

- Kingma and Welling (2013)
- Probabilistic (instead of deterministic)
  - Generative models via sampling from the latent space



VAE Loss Function

VAE Loss

- Cross Enthropy
  - **Reconstruction loss**
- Kullback-Leibler (KL) Divergence
  - Latent loss, divergence between Gaussian target and actual

$$D_{\mathrm{KL}}(P \parallel Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}$$

Anomaly Detection

Using AutoEncoders

- Unsupervised learning approach
- Build an auto-encoder to learn the representation of the null class
- Look at errors of the encoded model
  - **Large errors imply significant reconstruction losses: possible anomaly** (note: this can just be anomalous tails of the null class), however a good place to find real anomalies