TECHNISCHE UNIVERSITÄT
KAISERSLAUTERN

Artificial Neural Networks - Auto-Encoders

Sourajyoti Datta
Project – Collaborative Intelligence (DFKI)
Department of Computer Science
Technische Universität Kaiserslautern, Germany

Autoencoders

- **Artificial Neural Networks**
  - Trained to reconstruct it’s input in an unsupervised manner
  - Learns efficient data encodings
  - Generalization of Principal Component Analysis:
    - Learns a non-linear manifold

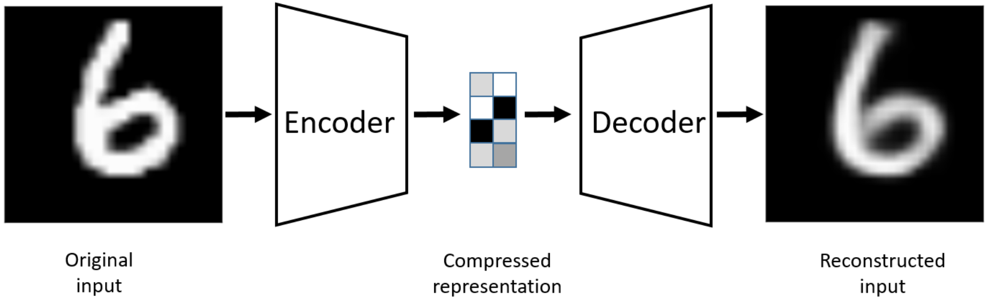


*Fig: Autoencoder example **

- **Tasks undertaken:**
  - A reduction network, that encodes the data
  - A reconstruction network, that generates the original information from the encoding

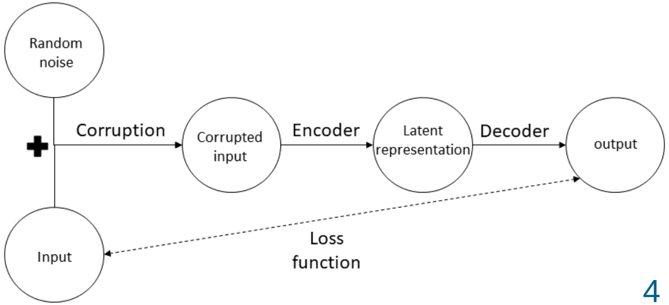
Types of Autoencoders

- **Regularized Autoencoders**
  - Encoders can simply learn the identity function
    - Given enough capacity of the encoder and the decoder, overfitting can occur (to the point where the network encodes input to an index)
  - Hence, the overfitting issue needs to be tackled
    - Traditionally, a ***bottleneck*** is imposed, which also provides low dimensional representation of the data. However, it can still cause overfitting.
  - Tackles the bias-variance tradeoff:
    - Reducing the reconstruction error vs. Generalizing the lower dimensional representation
- **Variational Autoencoders**
  - Generative models
    - Describes the generation of the data using probabilistic distributions.
    - Reflects the underlying causal relations, that have the potential for good generalization
  - Directed probabilistic graphical models (DPGM)
    - Whose posterior is approximated by a neural network, forming an autoencoder-like architecture

Regularized Autoencoders

- **Sparse Autoencoders**
  - Enforces sparsity on the hidden activation layers to deal with overfitting
    - Can be combined with bottleneck enforcement as well, of required
  - Similar to ordinary regularization, where they are applied on the activations instead of the weights
  - Two primary strategies:
    - *L1 Regularization*, which induces sparseness
    - *KL-Divergence,* which is a measure of the distance between two probability distributions

- **Denoising Autoencoders**
  - Can be either a regularization option, or a robust autoencoders for error correction
    - Input is disrupted by some noise
      - Using additive white Gaussian noise or Dropouts
    - Autoencoder is trained to reconstruct the clean version of the input



TECHNISCHE UNIVERSITÄT KAISERSLAUTERN

Regularized Autoencoders

- **Contractive Autoencoders**
  - Maps a neighborhood of input points to a smaller neighborhood of output points
  - Conditions the encoder be resistant to perturbations of the input
    - Emphasis on making the feature extraction less sensitive to small perturbations
    - Forces the encoder to disregard perturbations that are not important for reconstruction by the decoder
  - The regularizer corresponds to the L2-norm minimization of the Jacobian matrix of the network's activations with respect to the input
    - Penalty imposed on the Jacobian of the network, forces the model to learn useful information about the training distribution
    - The latent representations of the input tend to be similar, thus making reconstruction difficult
    - Variations in the latent representation not important for reconstruction would be diminished by the regularization, while important variations would remain due to their impact on reconstruction error

TECHNISCHE UNIVERSITÄT KAISERSLAUTERN

Variational Autoencoders

- VAEs are generative models that follow Variational Bayes (VB) Inference
  - Describe data generation through a probabilistic distribution
  - Equivalent to a probabilistic decoder
- A ***Reparameterization Trick*** is applied to estimate the variational lower bound
  - Results in an additional loss component and the Stochastic Gradient Variational Bayes (SGVB) estimator for the training algorithm

Advanced autoencoder techniques

- Autoencoders can suffer from low reconstruction quality (For e.g., blurry reconstructed images)
  - Based on the loss function
    - Does not account for realism of the result
    - For e.g., does not use the prior knowledge of the input images’ sharpness resulting in blurry output.
  - Hence, advanced techniques have been developed for the

- **Adversarially learned inference**
  - Generative Adversarial Networks (GANs)
    - The generator generates new samples
    - The discriminator distinguishes between real and generated samples
  - Suffers from mode collapse
    - Latent space represents only a part of the data, and drops modes from the distributions

Advanced autoencoder techniques

- **Deep feature consistent variational autoencoder**
  - Instead of measuring norm, a measure is used that also considers correlation
    - Instead of measuring the difference between the input/output directly, difference between their representation in the network layers is measured
    - Measuring difference at different layers imposes a more realistic difference measure for the autoencoder

- **Conditional image generation with PixelCNN decoders**
  - Another alternative proposes a composition between autoencoders and PixelCNN
    - Considers the local spatial statistics of the image
      - Using additive white Gaussian noise or Dropouts
    - Local statistics are replaced by the usage of an RNN, with the same concept in later developments

Applications of autoencoder

- **Generative Modelling**
  - VAEs are generative models that describe data generation through a probabilistic distribution
- **Classification**
  - Can be used in the semi-supervised setting for improving classification results
- **Clustering**
  - The latent representation serves as the input for any given clustering algorithm
- **Anomaly detection**
  - Follows the assumption that a trained autoencoder would learn the latent subspace of normal samples
  - Would result in a low reconstruction error for normal samples, and high reconstruction error for anomalies
- **Recommender Systems**
  - The latent representation serves as the input for Collaborative Filtering approaches
- **Dimensionality Reduction**
  - Learn a lower dimensional manifold based on the latent space structure