Autoencoders

R. Q. FEITOSA
Overview

Introduction

Autoencoders

Convolutional autoencoders

Denoising autoencoders

Variational autoencoders
Scarce labeled data

DL demands many labeled data

Lack of annotated RS data

Wealth of unlabeled RS data

Go unsupervised!
Overview

Introduction

Autoencoders

Convolutional autoencoders

Denoising autoencoders

Variational autoencoders
Autoencoders

Unsupervised

Dimensionality Reduction

Feature learning

A feedforward network

Same training machinery
AE vs PCA

Linear vs nonlinear dimensionality reduction

Source: https://www.edureka.co/blog/autoencoders-tutorial/
AE for dimensionality reduction

Aims at reproducing the input at the output. It comprises an encoder that contracts the data ... followed by a decoder that recovers the original dimension.
AE for dimensionality reduction

Aims at reproducing the input at the output.

The loss function relates to the similarity input ↔ output (e.g., $L_2$)

$$L(W) = \frac{1}{N} \sum_i \|x_i - \hat{x}_i\|^2 + \lambda R(W)$$
Actually, we are not interested in the output but in the “bottleneck that captures the most salient features of the training data.
Use of AE features as input to some ML classifier
Use of AE features as initialization for subsequent fine tuning
Overview

Introduction

Autoencoders

Convolutional autoencoders

Denoising autoencoders

Variational autoencoders
Convolutional Autoencoders

Aim at reproducing the input at the output
Overview

Introduction

Autoencoders

Convolutional autoencoders

Denoising autoencoders

Variational autoencoders
Denoising Autoencoders

Add noise to a clean image $x_i$ to create noisy versions ($\tilde{x}_{ij}$) of it.
Denoising Autoencoders

Train the autoencoder with the clean/noisy pairs \((x_i/\tilde{x}_{ij})\) to reconstruct the clean image from a noisy input.

Alain, Hinton, 2013). What regularized auto-encoders learn from the data generating distribute. ICLR~2013
Denoising Autoencoders

Apply the trained autoencoder to denoise any image.

Alain, Hinton, 2013). What regularized auto-encoders learn from the data generating distribution. ICLR~2013
Denoising Autoencoders: Example

Detection of Citrus Plantation Rows in UAV Images

La Rosa: Learning Geometric Features for Improving the Automatic Detection of Citrus Plantation Rows in UAV Images, 2020, (submitted to GRSL)
Denoising Autoencoders: Example

Detection of Citrus Plantation Rows in UAV Images

result of row detection algorithm before denoising
Denoising Autoencoders: Example

Detection of Citrus Plantation Rows in UAV Images

result of row detection algorithm after denoising
Denoising Autoencoders: Example

Detection of Citrus Plantation Rows in UAV Images
Speckel suppression in SAR Images

**Problem:** no clean image version for training the denoising AE

**Hypothesis:** no change between two consecutive dates.

Speckel suppression in SAR Images

Solution:

• Train a denoising AE with pairs of registered image patches of close dates, e.g., $x_{t1}$ to produce $x_{t2}$, and vice versa.

• Apply the trained autoencoder to overlapping patches of any date and build a mosaic with the central part of reconstructed patches.

From: Boulch, et al., 2018. Learning Speckle Suppression In SAR Images Without Ground Truth: Application To Sentinel-1 Time-series. IGARSS, 2018
Speckel suppression in SAR Images

Sample Result:

From: Boulch, et al., 2018. Learning Speckle Suppression In SAR Images Without Ground Truth: Application To Sentinel-1 Time-series. IGARSS, 2018
Overview

Introduction

Autoencoders

Convolutional autoencoders

Denoising autoencoders

Variational autoencoders
Recall GAN

- The GAN generator $G$ maps samples ($z$) drawn from a known distribution $q(z)$ to samples of an arbitrarily complex target distribution $p(x)$.

  $$G: Z \rightarrow X$$

- The GAN doesn’t tell which $z$ value will generate a given output (image) $x$.

  $$G^{-1}: X \rightarrow Z$$

Kingma and Welling, 2014, Auto-Encoding Variational Bayes
Graphical models

- Let’s assume that an image $x$ is determined by latent (non observable) variables $z$.
- Example: for face image, latent variables may refer to: pose, expression, hair style, gender, etc.
- Some methods (e.g., GAN) allow estimating $p_\theta(x|z)$.
- Estimating $p_\theta(z|x)$ is, however intractable $\rightarrow$ unknown.
VAE approach

- approximate $p_\theta(z|x)$ by another tractable distribution $q_\phi(z|x)$ (say, Gaussian), and

\[ q_\phi(z|x) \approx p_\theta(z|x) \]

- play around with its parameters in a way that $q_\phi(z|x)$ gets close enough to a collection of samples from $p_\theta(z|x)$. 
VAE Loss Function

... must enforce:

\[
\mathcal{L}(\theta, \phi) = -\mathbb{E}_{q_{\phi}(z|x)} \left[ \log p_{\theta}(x|z) \right] + KL \left( q_{\phi}(z|x) \parallel p_{\theta}(z) \right)
\]
The Kullback Leibler divergence

... measures how different are two probability distributions

Some definitions:

• **Information:**

\[ I = - \log p(x), \quad \text{where } x \text{ is an event} \]

Notice that, the lower the probability the higher is the information.

Ex.:

• “tomorrow will rain or not” → probability is 1, no information.
• “tomorrow temperature reaches 42 degrees” → probability is low, much information.
The Kullback Leibler divergence

... measures how different are two probability distributions

Some definitions:

- Entropy (or average information in a probability distribution)

\[ H = \mathbb{E}[-\log p(x)] = -\sum_x p(x) \log p(x) \]
The Kullback Leibler divergence

... measures how different are two probability distributions

Some definitions:

- **Kullback Leibler divergence**

\[
KL(p(x) \| q(x)) = - \sum_x p(x) \log q(x) + \sum_x p(x) \log p(x)
\]

\[
= \sum_x p(x) \log \frac{p(x)}{q(x)} = \mathbb{E}_{p(x)} \left[ \log \frac{p(x)}{q(x)} \right]
\]

Some properties

- \( KL(p(x) \| q(x)) \geq 0 \)
- \( KL(p(x) \| q(x)) \neq KL(q(x) \| p(x)) \) not a distance!
The Kullback Leibler divergence

... measures how different are two probability distributions

Definition:

\[
KL(p(x) || q(x)) = - \sum_x p(x) \log q(x) + \sum_x p(x) \log p(x) = \sum_x p(x) \log \frac{p(x)}{q(x)}
\]

Examples:

\[
KL(p(x) || q(x)) = 0
\]

\[
KL(p(x) || r(x)) = .5 \log \frac{.5}{.25} + .5 \log \frac{.5}{.75} = 0.0625
\]
Reconstruction error

The decoder maps $z$ to $\hat{x}$ in a deterministic way.

So, $p_\theta(x|z) \rightarrow p_\theta(x|\hat{x})$

- If $p_\theta$ is Gaussian it will have a term in this form
  $$p_\theta(x|\hat{x}) \approx \exp(-|x - \hat{x}|^2) \rightarrow$$
  The closer $x$ and $\hat{x}$ the larger is the term above.
- If $p_\theta$ is Bernoulli, it boils down to cross entropy.
**VAE Loss Function**

\[
\mathcal{L}(\theta, \phi) = -\mathbb{E}_{q_{\phi}(z|x)} \left[ \log p_{\theta}(x|z) \right] + KL \left( q_{\phi}(z|x) || p_{\theta}(z) \right)
\]

- **reconstruction error**
- **divergence between input-output** $q_{\phi}$ and $p_{\theta}$
How to train a VAE?

Instead of producing $z$, the encoder generates $\mu$ and $\Sigma$ of the target Gaussian distribution and samples from it.

**Problem:** Backprop cannot flow through a random node.
Reparameterization Trick

Introducing a variable \( \mathbf{\epsilon} \) allows us to reparameterize \( \mathbf{z} \) in a way that backprop flows through deterministic nodes.

\[ \mathbf{z} = \mathbf{\mu} + \text{diag}(\mathbf{\Sigma}) \odot \mathbf{\epsilon} \]

Improving realism

VAE and GAN combined to produce more realistic images. GAN’s Discriminator loss term in the VAE loss.

Source: Larsen et al. 2016 Autoencoding beyond pixels using a learned similarity metric
Disentangle latent variables

Check the region where images classes concentrate in the latent space.

By moving along the vector connecting two regions will cause a morphing from one to the other class.
Disentangle latent variables

Source: Larsen et al. 2016 Autoencoding beyond pixels using a learned similarity metric
Generating Face Images with VAE

Variational Autoencoder
Face Images Generation

hit here to watch the demo

Generating Face Images with VAE

Source: Steffen 2019 available at https://www.youtube.com/watch?v=uszj2MOLY08
VAE vs GAN

1. Both VAE and GAN are unsupervised learning

2. VAE
   - pro:
     - clear objective/cost function
     - Latent space more interpretable
   - con:
     - injected noise and imperfect reconstruction, result is blurred compared with GAN

3. GAN
   - pro:
     - result is better especially with noise? Nicer image.
   - con:
     - no clear object/cost function for comparison. Hard to train and converge
References on VAE

1. CARL DOERSCH, Tutorial on Variational Autoencoders,
Autoencoders

Thanks for your attention

R. Q. FEITOSA