Solving mode collapse with Autoencoder GANs

Mihaela Rosca

Thanks to: Balaji Lakshminarayanan, David Warde-Farley, Shakir Mohamed
Autoencoders

L₁/L₂ reconstruction loss

code
Adversarial autoencoders

Improve reconstruction quality by adding a GAN loss.
By construction, autoencoders learn to cover the entire training data.
Autoencoder GANs

Combine the reconstruction power of autoencoders with the sampling power of GANs!
How to sample?
Work on the code space
- not data space.
1) Learning the code distribution

Assumption: Learning the code distribution is simpler than learning the data distribution.
2) Match the code distribution to a desired prior

reconstructing

sampling
Working on the code space

Encoder

Decoder

Decoder

Decoder

learn p(codes)

match prior

reconstructing

sampling
Learning the code distribution
Learning the code distribution: PPGN
PPGN - key ingredients

- **Reconstructions**
  - Autoencoder + GAN + Perceptual loss in feature space

- **Samples**
  - Markov Chain
  - Conditioning
PPGN

L₁/L₂ reconstruction
GAN loss
Perceptual loss

Not depicted:
- Activation Maximization
- Encoder is pretrained classifier
PPGN: sampling using a denoising autoencoder

Learn $\frac{\partial \log p(c)}{\partial c}$ using a DAE, then sample using MCMC
Limitations of PPGN

Learning the code distribution

- Not end to end
  - Need to pretrain an encoder on the same dataset
- Depends on labels for:
  - Conditioning samples
  - Pretrained encoder
- Markov Chains
  - when to stop?
  - missing rejection step
Matching a desired prior
2) Match the code distribution to a desired prior
Sounds familiar?
Variational inference - the ELBO

\[ \log p_\theta(x) = \log \int p_\theta(x|z)p(z)dz \geq \mathbb{E}_{q_\eta(z|x)}[\log p_\theta(x|z)] - \text{KL}[q_\eta(z|x)||p(z)] \]
Variational autoencoders

\[ x \]

\[ q_\eta(z|x) \]

\[ \mathbb{E}_{q_\eta(z|x)}[\log p_\theta(x|z)] \]

\[ \text{computed analytically} \]

\[ \text{match prior} \]

\[ \text{computed analytically} \]

\[ \text{reconstructing} \]

\[ \text{sampling} \]

\[ \text{Decoder} \]

\[ \text{Decoder} \]

\[ p(z) \]

\[ \text{KL}[q_\eta(z|x) || p(z)] \]
**AlphaGAN**

combining GANs and VAEs

- variational inference
  - reconstructions
  - encoder network
- the posterior latent matches the prior

- implicit
  - encoder
  - decoder
- discriminators used to match distributions

*Variational Approaches for Auto-Encoding Generative Adversarial Networks*

M. Rosca, B. Lakshminarayanan, D. Warde-Farley, S. Mohamed
AlphaGAN

Estimated from samples:
- Data discriminator
- Codes discriminator

KL divergence:
$$KL[q_\eta(z|x)||p(z)]$$

Match prior:

Reconstructing:
$$E_{q_\eta(z|x)}[\log p_\theta(x|z)]$$

Sampling:

Encoder

Decoder
Q: how do we estimate the terms in the ELBO using GANs?
Density ratio trick

Estimate the ratio of two distributions only from samples, by building a binary classifier to distinguish between them.

\[
\frac{p(x)}{p_\theta(x)} = \frac{D(x)}{1 - D(x)}
\]
Using GANs for variational inference - the ELBO

\[
\log p_\theta(x) = \log \int p_\theta(x|z)p(z)dz \geq \mathbb{E}_{q_\eta(z|x)}[\log p_\theta(x|z)] - \text{KL}[q_\eta(z|x)||p(z)]
\]
ELBO - likelihood term

\[ \mathbb{E}_{q_\eta(z|x)}[\log p_\theta(x|z)] = \mathbb{E}_{q_\eta(z|x)}[\log \left( \frac{p_\theta(x|z)}{p(x)} p(x) \right)] \]

\[ = \mathbb{E}_{q_\eta(z|x)}[\log \frac{p_\theta(x|z)}{p(x)}] + \mathbb{E}_{q_\eta(z|x)}[\log p(x)] \]

\[ x \sim p(x) \]

\[ x \sim p_\theta(x|z) \]

Diagram:
- \( D \) with inputs \( x \sim p(x) \) and \( x \sim p_\theta(x|z) \) leading to \( 0/1 \) output.
ELBO - the KL term

\[- \text{KL}[q_\eta(z|x) \| p(z)] = \mathbb{E}_{q_\eta(z|x)} \left[ \log \frac{p(z)}{q_\eta(z|x)} \right] \]
From the ELBO to loss functions

We want to match:
- the reconstruction and data distributions
  - likelihood term
- the code and prior distributions
  - the KL

Tools for matching distributions:
- (GAN) the density ratio trick
- (VAE) observer likelihoods
  - reconstructions losses
New loss functions - via the density ratio trick

- **Reconstruction loss** (avoid mode collapse)
- **GAN loss** (improve recon quality)

**GAN loss** (make samples close to reconstructions)

**match prior**
Samples
Cifar10 - Inception score

Improved Techniques for Training GANs
T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen
CelebA - sample diversity

Improved Techniques for Training GANs
T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen
Matching the prior - AGE

- the encoder is the discriminator
  - forces data codes to match the prior
  - sample codes to not match the prior
Analysis of methods

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Observer</th>
<th>Likelihood Ratio estimator (&quot;synthetic&quot;)</th>
<th>KL (analytic)</th>
<th>KL (approximate)</th>
<th>Ratio estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAE</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DCGAN</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>VAE-GAN</td>
<td>✓</td>
<td>*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adversarial-VB</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>AGE</td>
<td>✓</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>$\alpha$-GAN (ours)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

Table 1: Comparison of different approaches for training generative latent variable models.
Summary

Try Autoencoder GANs if mode collapse is a problem.

Combining different learning principles results in a family of novel algorithms.
References

It Takes (Only) Two: Adversarial Generator-Encoder Networks
D. Ulyanov, A. Vedaldi, V. Lempitsky

Improved Techniques for Training GANs
T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, X. Chen

Variational Approaches for Auto-Encoding Generative Adversarial Networks
M. Rosca, B. Lakshminarayanan, D. Warde-Farley, S. Mohamed

Plug and play generative models
A. Nguyen, J. Clune, Y. Bengio, A. Dosovitskiy, J. Yosinski

Adversarial Autoencoders
A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, B. Frey

Adversarial Variational Bayes: Unifying Variational Autoencoders and Generative Adversarial Networks
L. Mescheder, S. Nowozin, A. Geiger

Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks
A. Radford, L. Metz, S. Chintala

Auto-Encoding Variational Bayes
D. P. Kingma, M. Welling

Generative Adversarial Networks
I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio