8. Other Deep Architectures

CS 519 Deep Learning, Winter 2017
Fuxin Li

With materials from Zsolt Kira and Ian Goodfellow
A brief overview of other architectures

• Unsupervised Architectures
  • Deep Belief Networks
  • Autoencoders
  • GANs

• Temporal Architectures
  • Recurrent Neural Networks (RNN)
  • LSTM

• We will carefully cover those items later
  • Right now just a brief overview in case that you might be tempted to use them in your project
Unsupervised Deep Learning

• CNN is most successful with a lot of training examples

• What can we do if we do not have any training example?
  • Or have very few of them?
Remember PCA: Characteristics and Limitations

- PCA works well when the data is near a linear manifold in high-dimensional space.
- Project the data onto this subspace spanned by principal components.
- In dimensions orthogonal to the subspace the data has low variance.
PCA as a “neural network”

- PCA goal:
  - Minimize reconstruction error
  
  \[
  \min_{\mathbf{V}} \sum_{i=1}^{n} (x_i - \mathbf{V}^\top \mathbf{V} x_i)^2
  \]
Generalize PCA to multi-layer nonlinear network

• Deep Autoencoder
  • Same as other NN (linear transform + nonlinearity + linear transform etc.)
  • Only difference is that after decoding, strive to reconstruct the original input
  • Can have convolutional/fully-connected/sparse versions
Krizhevsky’s deep autoencoder

The encoder has about 67,000,000 parameters.

It takes a few days on a GTX 285 GPU to train on two million images (Tiny dataset)

256-bit binary code

8192

4096

2048

1024

512

1024

1024

1024

2048

4096

8192

256-bit binary code

1024

1024

1024
Reconstructions of 32x32 color images from 256-bit codes
retrieved using 256 bit codes

retrieved using Euclidean distance in pixel intensity space
retrieved using 256 bit codes

retrieved using Euclidean distance in pixel intensity space
Restricted Boltzmann Machines

• Generative version of the encoder
• Binary-valued hidden variables
  • Define probabilities such as $P(h_i|X)$ and $P(x_i|H)$
  • You can generate samples of observed variables from hidden
• Think as an extension of probabilistic PCA
• Only if you are into generative models (PGM class)
• Unsupervised pre-training method to train it (Hinton, Salakhutdinov 2006)
• Convolutional and fully connected version available
Fooling a deep network (Szegedy et al. 2013)

- Optimizing a delta from the image to maximize a class prediction $f_c(x)$

$$\max_{\Delta I} f_c(I + \Delta I) - \lambda \|\Delta I\|^2$$

Goldfish (95.15% confidence)

Shark (93.89% confidence)

Giant Panda (99.32% confidence)

(Szegedy et al. 2013, Goodfellow et al. 2014, Nguyen et al. 2015)
Generative Adversarial Networks

$D(x)$ tries to be near 1

Differentiable function $D$

$x$ sampled from data

$D$ tries to make $D(G(z))$ near 0, $G$ tries to make $D(G(z))$ near 1

$x$ sampled from model

Differentiable function $G$

Input noise $z$
Generative Adversarial Networks

• Cost for the discriminator:
  • Standard cross-entropy loss, with everything from $p_{\text{data}}$ label 1, and everything from $z$ label 0

$$J^{(D)}(\theta^{(D)}, \theta^{(G)}) = -\frac{1}{2} \mathbb{E}_{x \sim p_{\text{data}}} \log D(x) - \frac{1}{2} \mathbb{E}_{z} \log (1 - D(G(z))).$$

• Cost for the generator:
  • Try to generate examples to “fool” the discriminator
DCGAN
Samples of DCGAN-generated images

Figure 18: Samples of images of bedrooms generated by a DCGAN trained on the LSUN dataset.
DCGAN representations
Text-to-Image with GANs

Figure 23: Text-to-image synthesis with GANs. Image reproduced from Reed et al. (2016b).
Text-to-Image with GANs

This small blue bird has a short pointy beak and brown on its wings.

This bird is completely red with black wings and pointy beak.

A small sized bird that has a cream belly and a short pointed bill.

A small bird with a black head and wings and features grey wings.
Problems

Figure 29: GANs on 128 × 128 ImageNet seem to have trouble with counting, often generating animals with the wrong number of body parts.
Problems

Figure 31: GANs on $128 \times 128$ ImageNet seem to have trouble coordinating global structure, for example, drawing “Fallout Cow,” an animal that has both quadrupedal and bipedal structure.
iGAN

https://www.youtube.com/watch?v=9c4z6YsBGQ0
Recurrent Neural Networks (RNNs)

- Temporal, Sequences
- Tied weights
- Some additional variants: Recursive Autoencoders, Long Short-Term Memory (LSTM)
Machine Translation

- Have to look at the entire sentence (or, many sentences)
Image Captioning

"man in black shirt is playing guitar."

"construction worker in orange safety vest is working on road."

"two young girls are playing with lego toy."

"boy is doing backflip on wakeboard."

"girl in pink dress is jumping in air."

"black and white dog jumps over bar."

"young girl in pink shirt is swinging on swing."

"man in blue wetsuit is surfing on wave."