Deep Learning Extensions

Lyle Ungar
University of Pennsylvania

Learning objectives
GANS
KL-divergence
CNNs for NLP
Semi-supervised learning

- Use unlabeled data $X_0$ to derive new features $z=\phi(x)$
- Train $y = f(\phi(x), w)$
Transfer learning

- Use one data set \((X_0, y_0)\) to train a model
- Find feature transformations \(\phi(x)\)
- Use those transformations \(\phi(x)\) on data from data set with a different label, \(y\).

Data set for learning \(\phi(x)\)

Data set with target \(y\).
Transfer learning for NNets

$X_0$ $\phi(x)$ $y_0$ Data set for learning $\phi(x)$

$X$ $\phi(x)$ $y$ Data set with target $y$. 

Digression: KL-divergence
Generative Adversarial Networks: GANS

\[
\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))]
\]

D: argmax \log(D(x)) + \log(1 - D(G(z)))

G: argmin \log(1 - D(G(z)))

https://medium.freecodecamp.org/an-intuitive-introduction-to-gans-7a2264a81394
GANs

- Make generative model, $G$, close to data PDF

$$\theta^* = \arg \min_{\theta} D_{KL} (p_{\text{data}}(x) \parallel p_{\text{model}}(x; \theta)) = \mathbb{E}(\sum p_{\text{data}} \log(p_{\text{model}}/p_{\text{data}}))$$

- For actual data this is the MLE:

$$\theta^* = \arg \max_{\theta} \prod_{i=1}^{m} p_{\text{model}}(x^{(i)}; \theta)$$

$$= \arg \max_{\theta} \log \prod_{i=1}^{m} p_{\text{model}}(x^{(i)}; \theta)$$

$$= \arg \max_{\theta} \sum_{i=1}^{m} \log p_{\text{model}}(x^{(i)}; \theta)$$

$$\mathbb{E}(\sum p_{\text{data}} \log(p_{\text{model}}/p_{\text{data}}))$$
GANs Training

- Training can be slow and unstable
  - Better convergence if you alternately train discriminator and generator
Conditional GANS for Image Translation (outline to photo)

\[
\min_{G} \max_{D} V(D, G) = \mathbb{E}_{x,y} [\log D(x, y)] + \\
\mathbb{E}_{x,z} [\log(1 - D(x, G(x, z)))]
\]

GANs: Image translation

Isola et al. (2016)

Isola et al. (2016)
GANs: Predict next video frame

Ground Truth  

MSE  

Adversarial

Lotter et al. 2015
GANs: Image translation

- Summer to winter
- Sketch to photo
- Low resolution to high resolution photo
- Young to old
- Photo to impressionist painting

https://www.youtube.com/watch?time_continue=3704&v=AJJRWFVfNPg
Visualizing networks

- Display pattern of hidden unit activations
  - Just shows they are sparse
- **Show** input that maximizes a node’s output
  - Over all inputs in the training set
  - Over the entire range of possible inputs
  - Early layers do feature detection
  - Later layers do object detection
- **Show** how occluding parts of an image affect classification accuracy

http://cs231n.github.io/understanding-cnn/
Maximally activating inputs for the first CONV layer of an AlexNet

http://cs231n.github.io/understanding-cnn/
Maximally activating images for some 5th maxpool layer neurons of an AlexNet.
P(correct label) after occlusion