● Learning Representations III
  ○ (Mixtures & GANs)
● Understanding Representations

CS331B: Representation Learning in Computer Vision
Amir R. Zamir
Silvio Savarese
Make-up class. Thursday, Hewlett 201, 5:30-7:00 PM.

Wednesday Class:
What we talked about so far...
I do not speak of what I have done. Such frivolous thoughts
rovect in my mind and now I am sorry to say that I have done the deed. I
have murdered the King of Scotland, King Duncan.

I have done the deed, and I am afraid that I shall sleep no more.

I have betrayed him in the most unmeaning way a person possibly could, and I’ve betrayed him twice, just like I have to Duncan,
where I have two as a dear friend. I wish that I could never do such a
performing thing, and I wish that I could sleep no more.

I was working anxiously for my lady to wound the hell that called
me in to do the deed. But before all else, I heard of the supernatural
repulsion, the thing that would be so well. But then my hand was in my
hand and I couldn’t grip it, but I could not sleep any more.

I have no matter to follow or discard it from my eye, but the other
creative created.

As I turned closer to Duncan’s room, I thought that I would panic
and freeze, for when I put on it, a sickening thought made me feel like I
was doing the right thing. As soon as Duncan fell I knew that it was
the best something can.

I heard him plinking as the dagger pierced through his skin,

I do not speak of what I have done. Such frivolous thoughts
rovect in my mind and now I am sorry to say that I have done the deed. I
have murdered the King of Scotland, King Duncan.

I have done the deed, and I am afraid that I shall sleep no more.

I have betrayed him in the most unmeaning way a person possibly could, and I’ve betrayed him twice, just like I have to Duncan,
where I have two as a dear friend. I wish that I could never do such a
performing thing, and I wish that I could sleep no more.

I was working anxiously for my lady to wound the hell that called
me in to do the deed. But before all else, I heard of the supernatural
repulsion, the thing that would be so well. But then my hand was in my
hand and I couldn’t grip it, but I could not sleep any more.

I have no matter to follow or discard it from my eye, but the other
creative created.

As I turned closer to Duncan’s room, I thought that I would panic
and freeze, for when I put on it, a sickening thought made me feel like I
was doing the right thing. As soon as Duncan fell I knew that it was
the best something can.

I heard him plinking as the dagger pierced through his skin.
Some basics concepts related to representations

- Ill-posedness
- Readout Non-linearity
- Dimensionality
- Computational Complexity
- Encoding power (i.e., performance)
- Narrowness of application domain (vertical vs horizontal representations)
Handcrafting Representations

Color Histograms

Deformable Part based Models (DPM)

Models based Shapes

Histogram of Gradients (HOG)

Felzenszwalb et al., 2010.
Dalal and Triggs, 2005.
Learning Representations

- **Supervised**
  - Representation constrained on task(s).

- **Unsupervised**
  - Representation constrained on reconstruction.
Unsupervised representation learning

- Sparse Coding
- Basic Autoencoders

New Image: $0.99 \times + 0.97 \times + 0.82 \times \ldots$
Supervised representation learning

Convolutional $X$

Neural Net

A Neuron
Supervised Low-level Matching

(a) true positives

(c) true negatives
Lectures 4&6

- 3D, activities, segmentations, layout, BoW, etc.
Objects-based Representations (ImageNet)

Query | Nearest Neighbors

Krizhevsky et al. 2012.
Deng et al. 2009.
Zeiler & Fergus. 2014.
Escorcia et al. 2015.
Scene-based Representations (MIT-Places)
(~static) Video Representations

Simonyan & Zisserman. 2014.
Karpathy et al. 2015.
Recurrent Models & Structured Prediction

Jain et al. 2016.
Today

- Mixing Representations
- Generative Adversarial Networks (GAN)
- Understanding and Probing Representations I
Methods of Mixing Representations
Mixing Representations

● Sometimes you wish to mix two/multiple tasks/representations
  ○ To transfer information or labeled data across tasks
  ○ To form a multi-task representation
  ○ To form a better single-task representation
Mixing Representations - How?

(a) Original Model

(test image) ---> ... ---> (old task 1) ---> ... ---> (old task m)
Mixing Representations - fine tuning

(a) Original Model

(b) Fine-tuning

Li & Hoiem. 2016.
Mixing Representations - joint (multi-task) training

(a) Original Model

Input:
(test image) $\theta_s$

Target:
(old task 1) $\theta_o$
(old task m)

(b) Fine-tuning

Input:
new task image

Target:
new task ground truth

(c) Joint Training

Input:
image for each task

Target:
old tasks’ ground truth
new task ground truth
Mixing Representations - feature extraction

(a) Original Model

Input: test image

Output: θ_s, θ_o

(b) Fine-tuning

Input: new task image

Target: new task ground truth

(c) Feature Extraction

Input: new task image

Target: new task ground truth

(d) Joint Training

Input: image for each task

Target: old tasks’ ground truth, new task ground truth

Li & Hoiem. 2016.
Mixing Representations - LwF

(a) Original Model

(b) Fine-tuning

(c) Feature Extraction

(d) Joint Training

(e) Learning without Forgetting

ECCV’16
Mixing Representations

- Pros and cons of each method:

<table>
<thead>
<tr>
<th></th>
<th>Fine Tuning</th>
<th>Duplicating and Fine Tuning</th>
<th>Feature Extraction</th>
<th>Joint Training</th>
</tr>
</thead>
<tbody>
<tr>
<td>new task performance</td>
<td>good</td>
<td>good</td>
<td>X medium</td>
<td>best</td>
</tr>
<tr>
<td>original task performance</td>
<td>X bad</td>
<td>good</td>
<td>good</td>
<td>good</td>
</tr>
<tr>
<td>training efficiency</td>
<td>fast</td>
<td>fast</td>
<td>fast</td>
<td>X slow</td>
</tr>
<tr>
<td>testing efficiency</td>
<td>fast</td>
<td>X slow</td>
<td>fast</td>
<td>fast</td>
</tr>
<tr>
<td>storage requirement</td>
<td>medium</td>
<td>X large</td>
<td>medium</td>
<td>X large</td>
</tr>
<tr>
<td>requires previous task data</td>
<td>no</td>
<td>no</td>
<td>no</td>
<td>X yes</td>
</tr>
</tbody>
</table>

Li & Hoiem. 2016.
Mixing Representations

- Empirical study:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>old</td>
<td>new</td>
<td>old</td>
<td>new</td>
<td>old</td>
<td>new</td>
<td>old</td>
</tr>
<tr>
<td>LwF (ours)</td>
<td>56.5</td>
<td>75.8</td>
<td>55.1</td>
<td>57.5</td>
<td>55.9</td>
<td>64.5</td>
<td>43.3</td>
</tr>
<tr>
<td>fine-tuning</td>
<td>-1.4</td>
<td>-0.3</td>
<td>-5.1</td>
<td>-1.5</td>
<td>-3.4</td>
<td>-1.0</td>
<td>-1.8</td>
</tr>
<tr>
<td>feat. extraction</td>
<td>0.5</td>
<td>-1.1</td>
<td>2.0</td>
<td>-5.3</td>
<td>1.2</td>
<td>-3.7</td>
<td>-0.2</td>
</tr>
<tr>
<td>joint training</td>
<td>0.2</td>
<td>0.0</td>
<td>0.5</td>
<td>-0.9</td>
<td>0.5</td>
<td>-0.6</td>
<td>-0.1</td>
</tr>
</tbody>
</table>
Curriculum Learning

For faster convergence, better minima, (and mixing representations)

Curriculum Learning

Guided learning helps training humans and animals

Start from simpler examples / easier tasks  (Piaget 1952, Skinner 1958)
The Dogma in question

It is best to learn from a training set of examples sampled from the same distribution as the test set. Really?
Question

Can machine learning algorithms benefit from a curriculum strategy?

Cognition journal:
(Elman 1993) vs (Rohde & Plaut 1999),
(Krueger & Dayan 2009)

Bengio et al. 2009. slides and paper credit.
Convex vs Non-Convex Criteria

- **Convex criteria**: the order of presentation of examples should not matter to the convergence point, but could influence *convergence speed*

- **Non-convex criteria**: the order and selection of examples could yield to a *better local minimum*
Deep Architectures

- Theoretical arguments: deep architectures can be exponentially more compact than shallow ones representing the same function
- Cognitive and neuroscience arguments
- Many local minima
- Good candidate for testing curriculum ideas

Deep Training Trajectories

(Random initialization)

(Unsupervised guidance)

(Erhan et al. AISTATS 09)

Bengio et al. 2009. slides and paper credit.
Starting from Easy Examples

1. Easiest
   • Lower level abstractions

2.

3. Most difficult examples
   • Higher level abstractions

Bengio et al. 2009. slides and paper credit.
Continuation Methods

- Target objective
- Heavily smoothed objective = surrogate criterion
- Track local minima
- Final solution
- Easy to find minimum

Bengio et al. 2009. slides and paper credit.
Curriculum Learning as Continuation

- Sequence of training distributions
- Initially peaking on easier / simpler ones
- Gradually give more weight to more difficult ones until reach target distribution

How to order examples?

- The right order is not known

- 3 series of experiments:
  1. Toy experiments with simple order
     - Larger margin first
     - Less noisy inputs first
  2. Simpler shapes first, more varied ones later
  3. Smaller vocabulary first

Bengio et al. 2009. slides and paper credit.
Larger Margin First: Faster Convergence

Easiness based on margin

Average test error

Input dimension

Bengio et al. 2009. slides and paper credit.
Cleaner First: Faster Convergence

Bengio et al. 2009. slides and paper credit.
Shape Recognition

First: easier, basic shapes

Second = target: more varied geometric shapes

Bengio et al. 2009. slides and paper credit.
Shape Recognition Experiment

- 3-hidden layers deep net known to involve local minima (unsupervised pre-training finds much better solutions)
- 10,000 training / 5,000 validation / 5,000 test examples

Procedure:
1. Train for k epochs on the easier shapes
2. Switch to target training set (more variations)

Shape Recognition Results

Bengio et al. 2009. slides and paper credit.
Why?

- Faster convergence to a minimum
- Wasting less time with noisy or harder to predict examples
- Convergence to better local minima

Curriculum = particular continuation method

- Finds better local minima of a non-convex training criterion
- Like a regularizer, with main effect on test set

Bengio et al. 2009. slides and paper credit.
Generative Adversarial Networks
This lecture

- Finishing up GANs
- Brief overview of representation understanding methods
- Generic Representations
Generative Adversarial Networks

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

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

Goodfellow et al. 2014.
Generative Adversarial Networks

Goodfellow et al. 2014.
Generative Adversarial Networks

- Why does this matter?

MNIST

CIFAR10

TFD

CIFAR100

Goodfellow et al. 2014.
Problems with GAN

- **Diversity**
  - Generator may overfit to certain modes in the data

- **Stability**
  - Hard to train. Sensitive to choice of hyperparameters and status of Discriminator & Generator

- **Evaluation**
  - How to evaluate the generated results?
Energy-Based GANs

- An energy-based formulation for discriminator
  - Auto-encoder instantiating

\[ f_D(x, z) = D(x) + [m - D(G(z))]^+ = \|\text{Dec}(\text{Enc}(x)) - x\| + [m - \|\text{Dec}(\text{Enc}(G(z))) - G(z)\|]^+ \]

\[ f_G(z) = \|D(G(z))\| = \|\text{Dec}(\text{Enc}(G(z))) - G(z)\| \]

Zhao et al. 2016.
Energy-Based GANs

- An energy-based formulation for discriminator
  - Auto-encoder instantiating
- Repelling Regularizer (Pull-Away). Minibatch Discrimination is an alternative.
- Better behaved, more diverse results.

\[ f_D(x, z) = D(x) + [m - D(G(z))]^+ \]
\[ = \|\text{Dec(Enc}(x)) - x\| + [m - \|\text{Dec(Enc}(G(z))) - G(z)\|]^+ \]

\[ f_G(z) = \|D(G(z))\| \]
\[ = \|\text{Dec(Enc}(G(z))) - G(z)\| \]

Zhao et al. 2016.
EBGAN vs GAN (on MNIST)

GAN

EBGAN

EBGAN with PT term

Zhao et al. 2016.
EBGAN vs GAN (on LSUN)

Zhao et al. 2016.
EBGAN vs GAN (on CelebA)

Zhao et al. 2016.
EBGAN (on ImageNet)

Zhao et al. 2016.
EBGAN (on ImageNet)

- Remember the main objective: learning an arbitrarily complex distribution of pixels (i.e. visual worlds)

Zhao et al. 2016.
A use case of such distribution: “Generative Visual Manipulation on the Natural Image Manifold”

- Editing images while remaining on the natural image manifold
  - i.e. preserving realism

- Learns distribution of real data using GAN. Defines a set of edits and constrain the output to fall on the manifold.

Zhu et al. 2016.
Zhu et al. 2016.
Zhu et al. 2016.
Understanding and Probing Representations

(very brief overview)
Understanding Representations

- Why?!
- Tools:
  - Nearest neighbors in full dimensional space
  - Low-dimensional embeddings
  - Read-out function
  - Inverting the representation (remember Hoggles?)
  - (Discussed in previous lectures:)
    - Minimal Image (what matters in an image)
    - Receptive field (what matters to a neuron)
    - Images maximally activating a neuron
    - Neuron activation maps
    - Visualization learned filters
Nearest neighbors in full dimensional space

Query | Nearest Neighbors

Krizhevsky et al. 2012.
Deng et al. 2009.
Nearest neighbors in full dimensional space

Zamir et al. 2016
Wang & Gupta. 2015
Agrawal et al. 2015
Krizhevsky (Imagenet), 2015
Low-dimensional embeddings

- 6000 MNIST Digits
  - tSNE
  - Isomap
  - Sammon M
  - LLE
Low-dimensional embeddings

- tSNE
Inverting the representation

Class appearance models (ImageNet)
Inverting the representation

- Supervised

- Dala & Triggs. 2005.
- Vondrick et al. 2013.
- Dosovitskiy & Brox. 2016.
Inverting the representation

- Supervised

Dosovitskiy & Brox

Mahendran & Vedaldi

AE

Dosovitskiy & Brox. 2016.
Hinton et al. 2006.
Understanding Representations

- Nearest neighbors in full dimensional space
- Low-dimensional embeddings
- Read-out function
- Inverting the representation
- Minimal Image (what matters in an image)
- Receptive field (what matters to a neuron)
- Images maximally activating a neuron
- Neuron activation maps
- Visualization learned filters
Course webpage:
http://web.stanford.edu/class/cs331b/
http://www.cs.stanford.edu/~amirz/
http://cvgl.stanford.edu/silvio/