Lecture 9: Introduction to Deep Learning & Convolutional Neural Networks

Alireza Akhavan Pour

CLASS.VISION
Image Classification: a core task in Computer Vision

(assume given set of discrete labels)
{dog, cat, truck, plane, ...}

→ cat
The problem: 
*semantic gap*

Images are represented as 3D arrays of numbers, with integers between [0, 255].

E.g.
300 x 100 x 3

(3 for 3 color channels RGB)
Challenges: Viewpoint Variation
Challenges: Illumination
Challenges: Deformation
Challenges: Occlusion
Challenges: Background clutter
Challenges: Intraclass variation
An image classifier

Unlike e.g. sorting a list of numbers,

no obvious way to hard-code the algorithm for recognizing a cat, or other classes.
Data-driven approach:

1. Collect a dataset of images and labels
2. Use Machine Learning to train an image classifier
3. Evaluate the classifier on a withheld set of test images

```python
def train(train_images, train_labels):
    # build a model for images -> labels...
    return model

def predict(model, test_images):
    # predict test_labels using the model...
    return test_labels
```
k-Nearest Neighbor on images *never used.*

- terrible performance at test time
- distance metrics on level of whole images can be very unintuitive

(All 3 images have same L2 distance to the one on the left)
Neural Networks & Deep Learning
number of arxiv papers submitted
Ng wants to quicken deep learning’s transition from research to real world application. Over the last five years deep learning has achieved huge success in energizing computer vision and speech recognition. Ng believes natural language processing is the next major field deep learning will revolutionize.

Ng foresees an explosion of job openings for engineers with AI expertise.

Ng is however concerned that a shortage of AI talents will stall its implementation in society, and so has pioneered online deep learning
Software 2.0

I sometimes see people refer to neural networks as just “another tool in your machine learning toolbox”. They have some pros and cons, they work here or there, and sometimes you can use them to win Kaggle competitions. Unfortunately, this interpretation completely misses the forest for the trees. Neural networks are not just another classifier, they represent the beginning of a fundamental shift in how we write software. They are Software 2.0.
Neural Networks practitioner
"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."
CNN

RNN

"straw"  "hat"  END

START  "straw"  "hat"

CNN\_\theta_c
Example with an image with 4 pixels, and 3 classes (cat/dog/ship)

![Input image](image)

<table>
<thead>
<tr>
<th></th>
<th>0.2</th>
<th>-0.5</th>
<th>0.1</th>
<th>2.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>W</td>
<td>1.5</td>
<td>1.3</td>
<td>2.1</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.25</td>
<td>0.2</td>
<td>-0.3</td>
</tr>
</tbody>
</table>

\[ f(x_i; W, b) = Wx_i + b \]

- Cat score: -96.8
- Dog score: 437.9
- Ship score: 61.95

Stretch pixels into single column.

\[
\begin{align*}
x_i &= \begin{pmatrix} 0.2 \\ 1.5 \\ 0.1 \\ 2.0 \\ 56 \\ 1.1 \\ 231 \\ 3.2 \\ 24 \\ -1.2 \\ 2 \\ 4 \end{pmatrix} \\
W &= \begin{pmatrix} 0.2 & -0.5 & 0.1 & 2.0 \\ 1.5 & 1.3 & 2.1 & 0.0 \\ 0 & 0.25 & 0.2 & -0.3 \\ 0.1 & 1 & 0.2 & 0 & 0 \\ 1.1 & 3.2 & 231 & 24 & 24 \\ 2.0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\end{pmatrix}
\]

\[ b &= \begin{pmatrix} 56 \\ 1.1 \\ 231 \\ 3.2 \\ 24 \\ -1.2 \\ 2 \\ 4 \end{pmatrix} \]
مقدمات شبکه عصبی:
ارقام دستنویس فارسی هدی
یک مدل ساده: طبقه‌بندی چند‌لایه

```python
model = Sequential()
model.add(Dense(64, activation='relu', input_dim=25))
model.add(Dense(10, activation='softmax'))
```

\[ softmax(L_n) = \frac{e^{L_n}}{\|e^L\|} \]
تعیین تابع هزینه و روش بهینه سازی شبکه

```python
model.compile(loss='categorical_crossentropy',
              optimizer='rmsprop',
              metrics=['accuracy'])
```

\[
- \sum Y'_i \cdot \log(Y_i)
\]

actural probabilities, “one-hot” encoded

\[
0.1 \ 0.2 \ 0.1 \ 0.3 \ 0.2 \ 0.1 \ 0.9 \ 0.2 \ 0.1 \ 0.1
\]

computed probabilities

\[
0.1 \ 0.2 \ 0.1 \ 0.3 \ 0.2 \ 0.1 \ 0.9 \ 0.2 \ 0.1 \ 0.1
\]
تعیین داده‌های آموزشی، سایز mini-batch و تعداد epoch

```
model.fit(x_train, y_train,
          epochs=30,
          batch_size=64)
```
تعیین داده‌های آموزشی، سایز mini-batch و تعداد epoch

```python
model.fit(x_train, y_train,
  epochs=30,
  batch_size=64)
```
تعیین داده‌های آموزشی، سایز mini-batch و تعداد epoch

```python
model.fit(x_train, y_train,
          epochs=30,
          batch_size=64)
```
تعیین داده‌های آموزشی، سایز mini-batch و تعداد epoch

```python
model.fit(x_train, y_train,
  epochs=30,
  batch_size=64)
```
تعیین داده‌های آموزشی، سایز mini-batch و تعداد epoch

```python
model.fit(x_train, y_train,
          epochs=30,
          batch_size=64)
```
Softmax
Cross-entropy
Mini-batch
Cross-entropy loss

- Training loss
- Test loss

Overfitting
Dropout
Main Idea: approximately combining exponentially many different neural network architectures efficiently

Overfitting ?!

Too many neurons

BAD Network

Not enough DATA
Convolutional layer
Fast-forward to today: ConvNets are everywhere

Classification

Retrieval

[Krizhevsky 2012]
Fast-forward to today: ConvNets are everywhere

Detection

Segmentation

[Faster R-CNN: Ren, He, Girshick, Sun 2015]

[Farabet et al., 2012]
Fast-forward to today: ConvNets are everywhere

self-driving cars

NVIDIA Tegra X1
Fast-forward to today: ConvNets are everywhere

[Simonyan et al. 2014]

[Goodfellow 2014]

[Taigman et al. 2014]
Fast-forward to today: ConvNets are everywhere

[Toshev, Szegedy 2014]

[Mnih 2013]
سه‌شنبه - ۸ تیم - ۷۹۳۱

SRTTU – A.Akhavan

Lecture 9
Fast-forward to today: ConvNets are everywhere

[Ciresan et al. 2013]

[Sermanet et al. 2011]

[Ciresan et al.]
Fast-forward to today: ConvNets are everywhere

[Denil et al. 2014]

[Turaga et al., 2010]
Whale recognition, Kaggle Challenge

Mnih and Hinton, 2010
Image Captioning

[Vinyals et al., 2015]
reddit.com/r/deepdream
سه‌شنبه - ۸۲ فروردین ۷۹۳۱
Deep Neural Networks Rival the Representation of Primate IT Cortex for Core Visual Object Recognition

[Cadieu et al., 2014]
ILSVRC Top-5 Error on ImageNet

- CV
- Deep Learning
- Human

Top-5 Error Rate (%)

- 2010
- 2011
- 2012
- 2013
- 2014
- Human
- 2015
- 2016
CONVNETS

HOW DO THEY WORK
Deep Learning on large images

64x64

1000x1000

Cat? (0/1)

1000x1000x3 = 3 million weights

3 million x 1000 = 3 billion weights
Computer Vision Problem

vertical edges

horizontal edges