Outline for April 15

- Recap Handout 16
- Other NN architectures
- What is unsupervised learning?
- First unsupervised learning algorithm: K-means clustering

- Lab 7 due TODAY (last chance for late day)
- Office hours TODAY 12:30-2pm
- Proposal due FRIDAY
Outline for April 15

• Recap Handout 16
• Other NN architectures
• What is unsupervised learning?
• First unsupervised learning algorithm: K-means clustering
Handout 16, Q1

(a) Which steps require parameter learning? (out of CONV, RELU, POOL, FLATTEN, FC)

(b) First layer params

(c) Second layer params

(d) Third layer params

(e) Total # params

(f) Compare to FC with 312,860?
(a) Which steps require parameter learning? (out of CONV, RELU, POOL, FLATTEN, FC)

(b) First layer params

(c) Second layer params

(d) Third layer params

(e) Total # params

(f) Compare to FC with 312,860?
Handout 16, Q1

(a) Which steps require parameter learning? (out of CONV, RELU, POOL, FLATTEN, FC) CONV, FC

(b) First layer params \(5 \times 5 \times 3 \times 20 + 20 = 1520\)

(c) Second layer params \(3 \times 3 \times 20 \times 10 + 10 = 1810\)

(d) Third layer params \(8 \times 8 \times 10 \times 10 + 10 = 6410\)

(e) Total # params \(9740\)

(f) Compare to FC with 312,860?
Handout 16, Q1

(a) Which steps require parameter learning? (out of CONV, RELU, POOL, FLATTEN, FC)

(b) First layer params
\[5 \times 5 \times 3 \times 20 + 20 = 1520\]

(c) Second layer params
\[3 \times 3 \times 20 \times 10 + 10 = 1810\]

(d) Third layer params
\[8 \times 8 \times 10 \times 10 + 10 = 6410\]

(e) Total # params
\[9740\]

(f) Compare to FC with 312,860?

Much better! (over order of magnitude better)
Handout 16, Q2

(a) $W=10$, $F=7$, $P=3$, $S=3$

(b) Draw padding

(c) Shade units where cross-correlation is performed
Handout 16, Q2

(a) $W=10$, $F=7$, $P=3$, $S=3$

\[
\frac{(10-7+6)}{3} + 1 = 4 \quad \text{(output size)}
\]

(b) Draw padding

(c) Shade units where cross-correlation is performed
Handout 16, Q2

(a) W=10, F=7, P=3, S=3

\[
(10-7+6)/3 + 1 = 4 \quad \text{(output size)}
\]

(b) Draw padding

(c) Shade units where cross-correlation is performed
(a) \(W=10, F=7, P=3, S=3\)

\[
\frac{(10-7+6)}{3} + 1 = 4 \quad \text{(output size)}
\]

(b) Draw padding

(c) Shade units where cross-correlation is performed
Handout 16, Q2

(a) $W=10$, $F=7$, $P=3$, $S=3$

$$\frac{10-7+6}{3} + 1 = 4 \quad \text{(output size)}$$

(b) Draw padding

(c) Shade units where cross-correlation is performed
(a) $W=10$, $F=7$, $P=3$, $S=3$

$$\frac{(10-7+6)}{3} + 1 = 4 \quad \text{(output size)}$$

(b) Draw padding

(c) Shade units where cross-correlation is performed
Handout 16, Q2

(a) $W=10, F=7, P=3, S=3$

\[
\frac{(10-7+6)}{3} + 1 = 4 \quad \text{(output size)}
\]

(b) Draw padding

(c) Shade units where cross-correlation is performed
Output size

\[ 32 = 1 + \frac{32 - 5 + 2P}{1} \]

\[ P = 2 \]

Size 5

Q3:

\[ W = 32 \]
\[ F = 5 \]
\[ S = 1 \]
\[ P = ? \]
Outline for April 15

• Recap Handout 16
• **Other NN architectures**
• What is unsupervised learning?
• First unsupervised learning algorithm: K-means clustering
A mostly complete chart of Neural Networks

©2016 Fjodor van Veen - asimovinstitute.org

- Backfed Input Cell
- Input Cell
- Noisy Input Cell
- Hidden Cell
- Probabilistic Hidden Cell
- Spiking Hidden Cell
- Output Cell
- Match Input Output Cell
- Recurrent Cell
- Memory Cell
- Different Memory Cell
- Kernel
- Convolution or Pool

Perceptron (P)
Feed Forward (FF)
Radial Basis Network (RBF)
Recurrent Neural Network (RNN)
Long / Short Term Memory (LSTM)
Gated Recurrent Unit (GRU)
Auto Encoder (AE)
Variational AE (VAE)
Denoising AE (DAE)
Sparse AE (SAE)

Image from Neural Network Zoo: http://www.asimovinstitute.org/neural-network-zoo/
Generative Adversarial Networks

- Idea: have two “adversarial” NNs
Generative Adversarial Networks

• Idea: have two “adversarial” NNs
• One network ("forger") is trying to produce new works in the style of its training data
Generative Adversarial Networks

- Idea: have two “adversarial” NNs
- One network (“forger”) is trying to produce new works in the style of its training data
- The other network (“art detective”) is trying to determine whether or not produced works are “fakes”
Generative Adversarial Networks

• Idea: have two “adversarial” NNs
• One network ("forger") is trying to produce new works in the style of its training data
• The other network ("art detective") is trying to determine whether or not produced works are "fakes"
• As training proceeds, the forger becomes better at producing works in a given style
Generative Adversarial Networks

- Idea: have two “adversarial” NNs
- One network (“forger”) is trying to produce new works in the style of its training data
- The other network (“art detective”) is trying to determine whether or not produced works are “fakes”
- As training proceeds, the forger becomes better at producing works in a given style
- But the art detective becomes better at discriminating between originals and fakes
GAN progress over time

4.5 years of GAN progress on face generation. arxiv.org/abs/1406.2661
arxiv.org/abs/1511.06434
arxiv.org/abs/1606.07536
arxiv.org/abs/1710.10196
arxiv.org/abs/1812.04948
GAN painting

 Sold for almost half a million dollars
Outline for April 15

• Recap Handout 16
• Other NN architectures
• What is unsupervised learning?
• First unsupervised learning algorithm: K-means clustering
Supervised Learning: makes use of examples where we know the underlying “truth” (label/output)

Unsupervised Learning: Learn underlying structure or features without labeled training data
Unsupervised learning: 3 main areas

1) **Clustering**: group data points into clusters based on features only

2) **Dimensionality reduction**: remove feature correlation, compress data, visualize data

3) **Structured prediction**: model latent variables (example: Hidden Markov Models)
Unsupervised learning examples from biology: clustering

<table>
<thead>
<tr>
<th>$K = 2$</th>
<th>30 runs</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K = 3$</td>
<td>9 runs</td>
</tr>
<tr>
<td>$K = 4$</td>
<td>30 runs</td>
</tr>
<tr>
<td>$K = 5$</td>
<td>30 runs</td>
</tr>
<tr>
<td>$K = 6$</td>
<td>23 runs</td>
</tr>
<tr>
<td>$K = 7$</td>
<td>7 runs</td>
</tr>
<tr>
<td>$K = 7$</td>
<td>2 runs</td>
</tr>
</tbody>
</table>

Figure: German Dziebel
Unsupervised learning examples from biology: structured prediction
Unsupervised learning examples from biology: structured prediction

The complete genome sequence of a Neanderthal from the Altai Mountains, Prufer et al (2014)
Outline for April 15

• Recap Handout 16
• Other NN architectures
• What is unsupervised learning?
• First unsupervised learning algorithm: K-means clustering
**K-means**

Goal: find $K$ clusters $\{C_1, C_2, \ldots, C_K\}$ with means $\mu_1, \mu_2, \ldots, \mu_K$ in $\mathbb{R}^p$

Minimizes

$$J(C) = \sum_{k=1}^{K} \sum_{x \in C_k} ||x - \mu_k||^2$$

Within cluster sum of squares

Red distances = within cluster sum of squares
Algorithm

Initialization
Choose $\hat{\mu}_1^{(0)}, \hat{\mu}_2^{(0)}, \ldots, \hat{\mu}_k^{(0)}$
(usually points from training data)

E-step
Assign each point $\tilde{x}$ a cluster $k$ based on the min distance to each cluster mean $\hat{\mu}_k^{(t)}$

$\Rightarrow \tilde{x} \in C_k^{(t)}$

M-step
Update cluster means

$\hat{\mu}_k^{(t+1)} = \frac{1}{|\mathcal{E}_k^{(t)}|} \sum_{i \in \mathcal{E}_k^{(t)}} \tilde{x}_i$

Stop
* no cluster membership changes (means don't change)
* max iter exceeded.
* see config you saw before (cycle)
Example with $p=2$, $n=16$

Note: normally we would choose initial means from among training data point, but here we did not.
Example with $p=2$, $n=16$
1. runtime? $K, n, p$
2. how to choose $K$?

NEXT TIME!