Today:
- Learning representations III
  - Deep Belief Networks
  - ICA
  - CCA
    - Neuroscience example
    - Latent Dirichlet Allocation

Readings:
- 

Deep Belief Networks
[Hinton & Salakhutdinov, Science, 2006]

- Problem: training networks with many hidden layers doesn’t work very well
  - local minima, very slow training if initialize with zero weights

- Deep belief networks
  - autoencoder networks to learn low dimensional encodings

  - but more layers, to learn better encodings
Deep Belief Networks
[Hinton & Salakhutdinov, 2006]

Original image reconstructed from 2000-1000-500-30 DBN
reconstructed from 2000-300, linear PCA

versus

logistic transformations

linear transformations

Encoding of digit images in two dimensions
[Hinton & Salakhutdinov, 2006]

784-2 linear encoding (PCA) 784-1000-500-250-2 DBNet
Restricted Boltzman Machine

- Bipartite graph, logistic activation
- Inference: fill in any nodes, estimate other nodes
- consider $v_i, h_j$ are boolean variables

\[
P(h_j = 1|v) = \frac{1}{1 + \exp(\sum_i w_{ij}v_i)}
\]
\[
P(v_i = 1|h) = \frac{1}{1 + \exp(\sum_j w_{ij}h_j)}
\]

Deep Belief Networks: Training

Fig. 1. Pretraining consists of learning a stack of restricted Boltzmann machines (RBMs), each having only one layer of feature detectors. The learned feature activations of one RBM are used as the “data” for training the next RBM in the stack. After the pretraining, the RBMs are “unrolled” to create a deep autoencoder, which is then fine-tuned using backpropagation of error derivatives.

[Hinton & Salakhutdinov, 2006]
Independent Components Analysis (ICA)

- PCA seeks orthogonal directions \( \langle Y_1 \ldots Y_M \rangle \) in feature space \( X \) that minimize reconstruction error.

- ICA seeks directions \( \langle Y_1 \ldots Y_M \rangle \) that are most statistically independent. I.e., that minimize \( I(Y) \), the mutual information between the \( Y_j \):
  \[
  I(Y) = \sum_{j=1}^{J} H(Y_j) - H(Y)
  \]

  \[x \rightarrow \text{ICA} \rightarrow x\]

Dimensionality reduction across multiple datasets

- Given data sets A and B, find linear projections of each into a common lower dimensional space!
  - Generalized SVD: minimize square reconstruction errors of both.
  - Canonical correlation analysis: maximize correlation of A and B in the projected space.

  \[\text{learned shared representation}\]

  \[\text{data set A} \quad \text{data set B}\]
An Example Use of CCA

arbitrary word → \textbf{Generative theory} of word representation → predicted brain activity
fMRI activation for "bottle":

Mean activation averaged over 60 different stimuli:

"bottle" minus mean activation:

Idea: Predict neural activity from corpus statistics of stimulus word

[Mitchell et al., Science, 2008]
<table>
<thead>
<tr>
<th>Semantic feature values:</th>
<th>“celery”</th>
<th>Semantic feature values:</th>
<th>“airplane”</th>
</tr>
</thead>
<tbody>
<tr>
<td>“celery”</td>
<td>0.8368, eat</td>
<td>“airplane”</td>
<td>0.8673, ride</td>
</tr>
<tr>
<td>0.3461, taste</td>
<td></td>
<td>0.2891, see</td>
<td>0.2851, say</td>
</tr>
<tr>
<td>0.3153, fill</td>
<td></td>
<td>0.1689, near</td>
<td>0.1228, open</td>
</tr>
<tr>
<td>0.2430, see</td>
<td></td>
<td>0.0883, hear</td>
<td>0.0771, run</td>
</tr>
<tr>
<td>0.1145, clean</td>
<td></td>
<td>0.0749, lift</td>
<td>0.0049, smell</td>
</tr>
<tr>
<td>0.0600, open</td>
<td></td>
<td></td>
<td>0.0010, wear</td>
</tr>
<tr>
<td>0.0586, smell</td>
<td></td>
<td></td>
<td>0.0000, taste</td>
</tr>
<tr>
<td>0.0286, touch</td>
<td></td>
<td></td>
<td>0.0000, rub</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td>0.0000, manipulate</td>
</tr>
<tr>
<td>0.0000, ride</td>
<td></td>
<td></td>
<td>0.0000, manipulate</td>
</tr>
</tbody>
</table>

**Predicted Activation is Sum of Feature Contributions**

\[
prediction_v = \sum_{i=1}^{25} f_i(w) c_{141382, eat}^{vi}
\]

500,000 learned parameters

Predicted “Celery”
Predicted and observed fMRI images for “celery” and “airplane” after training on 58 other words.

Evaluating the Computational Model

- **Train** it using 58 of the 60 word stimuli
- **Apply** it to predict fMRI images for other 2 words
- **Test**: show it the observed images for the 2 held-out, and make it predict which is which

1770 test pairs in leave-2-out:
- Random guessing $\rightarrow$ 0.50 accuracy
- Accuracy above 0.61 is significant ($p<0.05$)

*Mean accuracy over 9 subjects*: 0.79
Q4: What are the **actual** semantic primitives from which neural encodings are composed?

![Diagram showing word to neural representation flow]

Alternative semantic feature sets

<table>
<thead>
<tr>
<th>Feature Set</th>
<th>Mean Acc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>PREDEFINED corpus features</td>
<td></td>
</tr>
<tr>
<td>25 verb co-occurrences</td>
<td>.79</td>
</tr>
<tr>
<td>486 verb co-occurrences</td>
<td>.79</td>
</tr>
<tr>
<td>50,000 word co-occurrences</td>
<td>.76</td>
</tr>
<tr>
<td>300 Latent Semantic Analysis features</td>
<td>.73</td>
</tr>
<tr>
<td>50 corpus features from Collobert&amp;Weston ICML08</td>
<td>.78</td>
</tr>
<tr>
<td>218 features collected using Mechanical Turk*</td>
<td>.83</td>
</tr>
<tr>
<td>20 features discovered from the data**</td>
<td>.87</td>
</tr>
</tbody>
</table>

* developed by Dean Pommerleau
** developed by Indra Rustandi
Discovering shared semantic basis

[Rustandi et al., 2009]

\[ \text{predict representation } v = \sum f(w) c_u \]

**specific to study/subject**

\[ \text{predict representation } v = \sum f(w) c_u \]

\[ \text{predict representation } v = \sum f(w) c_u \]

\[ \text{predict representation } v = \sum f(w) c_u \]

\[ \text{predict representation } v = \sum f(w) c_u \]

\[ \text{predict representation } v = \sum f(w) c_u \]

* trained using Canonical Correlation Analysis

---

**Multi-study (WP+WO) Multi-subject (9+11) CCA**

**Top Stimulus Words**

<table>
<thead>
<tr>
<th>component 1</th>
<th>component 2</th>
<th>component 3</th>
<th>component 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>most active stimuli</td>
<td>apartment</td>
<td>church</td>
<td>screwdriver</td>
</tr>
<tr>
<td></td>
<td>closet</td>
<td>pliers</td>
<td>refrigerator</td>
</tr>
<tr>
<td></td>
<td>house</td>
<td>knife</td>
<td>butter fly</td>
</tr>
<tr>
<td></td>
<td>barn</td>
<td>hammer</td>
<td>bicycle</td>
</tr>
</tbody>
</table>

shelter? manipulation? things that touch me?