Unsupervised Learning from Data

In Machine Translation

Unsupervised methods are used to go from sentence pairs to word pairs

<table>
<thead>
<tr>
<th>French</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>la maison est grand</td>
<td>the house is big</td>
</tr>
<tr>
<td>c'est un haricot vert</td>
<td>it's a green bean</td>
</tr>
<tr>
<td>je le lui donne</td>
<td>I give it to him</td>
</tr>
</tbody>
</table>

⇒

\[
\begin{align*}
  p(la|the) & = 0.453 \\
  p(le|the) & = 0.334 \\
  p(maison|house) & = 0.876 \\
  p(bleu|blue) & = 0.963
\end{align*}
\]

Done using EM Algorithm

- Incomplete data
  - if we had complete data, we could estimate model
  - if we had model, we could fill in the gaps in the data
- Expectation Maximization (EM) in a nutshell
  1. initialize model parameters (e.g. uniform)
  2. assign probabilities to the missing data
  3. estimate model parameters from completed data
  4. iterate steps 2–3 until convergence

Possibly other attempts to infer from Big Data

Use Unsupervised Machine Learning?
Learning Word Translations

- Initial step: all alignments equally likely

- After one iteration
  - Alignments, e.g., between `la` and `the` are more likely

- After another iteration
  - It becomes apparent that alignments, e.g., between `fleur` and `flower` are more likely (pigeon hole principle)

- Eventually converges, inherent hidden structure found by EM
  - Learns translation probabilities

\[
\begin{align*}
  p(\text{la}|\text{the}) & \approx 0.453 \\
  p(\text{le}|\text{the}) & \approx 0.334 \\
  p(\text{maison}|\text{house}) & \approx 0.876 \\
  p(\text{bleu}|\text{blue}) & \approx 0.563
\end{align*}
\]
Documents exhibit multiple topics: biology, genetics, computation, ...
Hidden Markov Models for Speech or Actions

- words/action visible as an evidence sequence
- can be modeled with a HMM with transitions between states and visible evidence for a state

Learning HMM from Data

- Possible to use unsupervised learning to find probabilities concerning hidden variables from data with just visible evidence
- Used in Speech Recognizers, Activity Recognizers