A Virtual Manipulative for Learning Log-Linear Models

Francis Ferraro, and Jason Eisner
http://cs.jhu.edu/~jason/tutorials/loglin
Log-Linear Modeling, Or How I Learned to Stop Worrying and Love (and Use) MaxEnt
Manipulatives

Cuisenaire rods
Arithmetic concepts

Geoboards
Geometric concepts

http://commons.m.wikimedia.org/wiki/File:Cuisenaire_staircase.JPG
http://en.m.wikipedia.org/wiki/File:Geoboard.JPG
Virtual Manipulatives

- National Library of Virtual Manipulatives (NLVM); see Clements and McMillen (1996), and Moyer (2002)

- In NLP:
  - Parsing board game (van Halteren, 2002)
  - HMMs (Eisner, 2002)
  - Core NLP technology web interface (Light et al., 2005)
  - Model 1 stack decoder (github.com/mjpost/stack-decoder)

- Speech and ML: see paper
Online Tutorial

- http://cs.jhu.edu/~jason/tutorials/loglin/
  - Chrome, Firefox, Safari, IE 10+
  - Standard computers (touchscreens not fully supported yet)
- Open source
Summary of Materials

- 18 lessons, including conclusion/summary
- Auxiliary handout with some derivation details
- "Additional resources" page with...
  - links to further readings, tutorials
  - code, toolkit pointers
  - additional materials (coming soon)

- We welcome suggestions
Student Take-Aways

- Internalize underlying log-linear concepts
- Familiar objects to manipulate
- Learn by playing
Student Take-Aways

- Internalize underlying log-linear concepts
- Familiar objects to manipulate
- Learn by playing

\[ p_{\theta}(y \mid x) \propto \exp \left( \bar{\theta} \cdot \bar{f}(x, y) \right) \]
Student Take-Aways

- Internalize underlying log-linear concepts
- Familiar objects to manipulate
- Learn by playing

\[ p_{\theta}(y \mid x) \propto \exp \left( \theta \cdot f(x, y) \right) \]

\[ F(\theta) = \left( \sum_{i=1}^{N} \log p_{\theta}(y_i \mid x_i) \right) - C \cdot R(\theta) \]
Student Take-Aways

- Internalize underlying log-linear concepts
- Familiar objects to manipulate
- Learn by playing

\[
p_{\theta}(y \mid x) \propto \exp \left( \theta \cdot f(x, y) \right)
\]

\[
F(\theta) = \left( \sum_{i=1}^{N} \log p_{\theta}(y_i \mid x_i) \right) - C \cdot R(\theta)
\]

\[
\mathbb{E}_{\tilde{p}} \left[ \tilde{f}(X, Y) \right] - \mathbb{E}_{p_{\theta}} \left[ f(X, Y) \right]
\]
Student Take-Aways

\[ p_\theta(y \mid x) \]

Choosing an appropriate model
Student Take-Aways

\[ p_\theta(y \mid x) \]

Choosing an appropriate model

\[ \text{Unconditioned vs. conditional models} \]
Student Take-Aways

$p_\theta(y \mid x)$

Choosing an appropriate model

$\tilde{f}(X,Y)$

Feature design

$P(\bigcirc)$

Unconditioned vs. conditional models
Student Take-Aways

\[ p_{\theta}(y \mid x) \]

Choosing an appropriate model

\[ \mathbb{E}_{p_{\theta}} \left[ f(X, Y) \right] \]

Feature counts

\[ \mathbb{P}(\bullet \mid \bullet) \]

Unconditioned vs. conditional models
Student Take-Aways

$p_{\theta}(y \mid x)$

Choosing an appropriate model

$\mathbf{f}(X, Y)$

Feature design

$\mathbb{E}_{p_{\theta}} \left[ \mathbf{f}(X, Y) \right]$  

Feature counts

$P(\bigcirc \mid \Box)$

Unconditioned vs. conditional models

$P(\Box)$

Generalization
Student Take-Aways

\( p_\theta(y \mid x) \)

Choosing an appropriate model

\( \mathbb{E}_{p_\theta} \left[ \tilde{f}(X, Y) \right] \)

Feature counts

\( \hat{\theta} \)

Statistical estimation
Student Take-Aways

\[ p_\theta(y \mid x) \]

Choosing an appropriate model

\[ \mathbb{E}_{p_\theta} \left[ \tilde{f}(X, Y) \right] \]

Feature counts

\[ \hat{\theta} \]

Statistical estimation
Student Take-Aways

\[ p_\theta(y \mid x) \]
Choosing an appropriate model

\[ f(X, Y) \]
Feature design

\[ P(\bullet \mid \bullet) \]
Unconditioned vs. conditional models

\[ \mathbb{E}_{p_\theta} \left[ f(X, Y) \right] \]
Feature counts

\[ P(\bullet \mid \bullet \bullet) \]
Generalization

\[ \hat{\theta} \]
Statistical estimation
Student Take-Aways

$p_{\theta}(y \mid x)$

Choosing an appropriate model

$\bar{f}(X, Y)$

Feature design

$P(\bullet \mid \circ)$

Unconditioned vs. conditional models

$\mathbb{E}_{p_{\theta}} \left[ \bar{f}(X, Y) \right]$  

Feature counts

$P(\circ \mid \circ)$

Generalization

$\hat{\theta}$

Statistical estimation
Student Take-Aways

\[ p_\theta(y | x) \]

Choosing an appropriate model

\[ \mathbb{E}_{p_\theta} \left[ \tilde{f}(X, Y) \right] \]

Feature counts

\[ P(\bigcirc | \bigdiamond) \]

Unconditioned vs. conditional models

\[ P(\bigdiamond) \]

Feature design

\[ P(\bigstar | \bigdiamond) \]

Generalization

\[ \hat{\theta} \]

Statistical estimation
Our Experiences

- Users with and without CS/NLP interests
  - JHU Fall 2012 NLP class
  - LSA Summer 2013 Intro to Computational Linguistics
  - Others with no technical/linguistic background
Our Experiences

- Users with and without CS/NLP interests
  - JHU Fall 2012 NLP class
  - LSA Summer 2013 Intro to Computational Linguistics
  - Others with no technical/linguistic background
Use in Your Class, too!

- Link to ours:
  http://cs.jhu.edu/~jason/tutorials/loglin
Use in Your Class, too!

- Link to ours:
  http://cs.jhu.edu/~jason/tutorials/loglin

- ... or set up your own
  - Nothing to install
  - Extend/modify lessons with simple text file
Thanks!

- http://cs.jhu.edu/~jason/tutorials/loglin

- Questions?
- Suggestions?
- Comments?
Available Shapes