Cognitive Models of Test-Item Effects in Human Category Learning

Xiaojin Zhu, Bryan R. Gibson, Kwang-Sung Jun,
Timothy T. Rogers*, Joseph Harrison*, Chuck Kalish†

Departments of Computer Sciences, Psychology*, & Educational Psychology†
University of Wisconsin-Madison

ICML 2010

A Typical Human Category Learning Experiment

training:

stimulus $\mathbf{x}$

A Typical Human Category Learning Experiment

training: key

stimulus $\mathbf{x}$ $f(\mathbf{x})$

A Typical Human Category Learning Experiment

training:

key

stimulus $\mathbf{x}$ $f(\mathbf{x})$ feedback y

A Typical Human Category Learning Experiment

training:  key  yes 

stimulus $\mathbf{x}$ $f(\mathbf{x})$ feedback y

A Typical Human Category Learning Experiment

training:  key  yes  key

stimulus $\mathbf{x}$ $f(\mathbf{x})$ feedback y

A Typical Human Category Learning Experiment

training:  key  yes  key  no

stimulus $\mathbf{x}$ $f(\mathbf{x})$ feedback y

A Typical Human Category Learning Experiment

training:  key  yes  key  no . . .

stimulus $\mathbf{x}$ $f(\mathbf{x})$ feedback y

A Typical Human Category Learning Experiment

training:  key  yes  key  no . . .

stimulus $\mathbf{x}$ $f(\mathbf{x})$ feedback y

A Typical Human Category Learning Experiment

training:  key  yes  key  no . . .

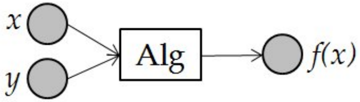
stimulus $\mathbf{x}$ $f(\mathbf{x})$ feedback y

test:  key  key . . .

One Goal of Cognitive Psychology

... is to identify the algorithm in our mind

CogSci	Machine Learning
stimulus	feature vector $\mathbf{x}$
category feedback	class y
stimulus with feedback	labeled data $(\mathbf{x}, y)$
stimulus without feedback	unlabeled data $\mathbf{x}$
response	classification $f(\mathbf{x})$



Human Semi-Supervised Learning?

training:  key  yes  key  no . . .

stimulus $\mathbf{x}$ feedback

- A computer can hold a trained classifier f fixed during testing.
- A human may not

test:  key  key . . .

Test-Item Effect

This work answers two questions:

1. Will unlabeled test items change the classifier in humans mind?

Test-Item Effect

This work answers two questions:

1. Will unlabeled test items change the classifier in humans mind? (yes)

Test-Item Effect

This work answers two questions:

1. Will unlabeled test items change the classifier in humans mind? (yes)
    - Two identical people A, B receiving exactly the same training data

Test-Item Effect

This work answers two questions:

1. Will unlabeled test items change the classifier in humans mind? (yes)
    - Two identical people A, B receiving exactly the same training data
    - The test data (without label feedback) is different

Test-Item Effect

This work answers two questions:

1. Will unlabeled test items change the classifier in humans mind? (yes)
   - Two identical people A, B receiving exactly the same training data
   - The test data (without label feedback) is different
   - Because of this difference, they disagree on certain test items

Test-Item Effect

This work answers two questions:

1. Will unlabeled test items change the classifier in humans mind? (yes)
   - Two identical people A, B receiving exactly the same training data
   - The test data (without label feedback) is different
   - Because of this difference, they disagree on certain test items

Test-Item Effect

This work answers two questions:

1. Will unlabeled test items change the classifier in humans mind? (yes)
   - Two identical people A, B receiving exactly the same training data
   - The test data (without label feedback) is different
   - Because of this difference, they disagree on certain test items

2. How to model test-item effect?

Test-Item Effect

This work answers two questions:

1. Will unlabeled test items change the classifier in humans mind? (yes)
   - Two identical people A, B receiving exactly the same training data
   - The test data (without label feedback) is different
   - Because of this difference, they disagree on certain test items

2. How to model test-item effect? (3 semi-supervised models)

Test-Item Effect 1: Order of Test Items

- 1D feature space 

Test-Item Effect 1: Order of Test Items

- 1D feature space 
- 10 labeled items, five pairs of $(\mathbf{x}, y) = (-2, 0), (2, 1)$

Test-Item Effect 1: Order of Test Items

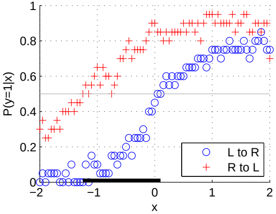
- 1D feature space

- 10 labeled items, five pairs of $(\mathbf{x}, y) = (-2, 0), (2, 1)$
- Two conditions, 20 subjects each:
  - **L to R**: test item -2,-1.95,-1.9, . . . , 2
  - **R to L**: reverse order.

Test-Item Effect 1: Order of Test Items

- 1D feature space 

- 10 labeled items, five pairs of $(\mathbf{x}, y) = (-2, 0), (2, 1)$
- Two conditions, 20 subjects each:
  - **L to R**: test item -2,-1.95,-1.9, . . . , 2
  - **R to L**: reverse order.

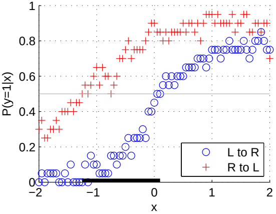


- Subjects in “L to R” classify more test items as $y = 0$, and vice versa.

Test-Item Effect 1: Order of Test Items

- 1D feature space

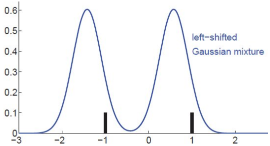
- 10 labeled items, five pairs of $(\mathbf{x}, y) = (-2, 0), (2, 1)$
- Two conditions, 20 subjects each:
  - **L to R**: test item -2,-1.95,-1.9, . . . , 2
  - **R to L**: reverse order.



- Subjects in "L to R" classify more test items as $y = 0$, and vice versa.
- For test items in [-1.2, 0.1], a majority-vote among subjects will classify them in opposite ways in these two conditions.

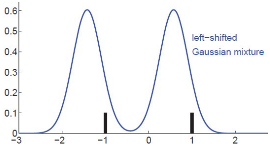
Test-Item Effect 2: Distribution of Test Items [AAAI 07]

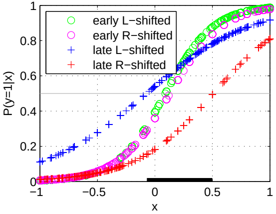
- Same feature space
- 20 labeled items, ten pairs of $(\mathbf{x}, y) = (-1, 0), (1, 1)$
- 22 subjects. Test items drawn from two-component GMM. Two conditions:
  - **L shifted**: GMM $\mu_1 = -1.43, \mu_2 = 0.57$
  - **R shifted**: GMM $\mu_1 = -0.57, \mu_2 = 1.43$



Test-Item Effect 2: Distribution of Test Items [AAAI 07]

- Same feature space
- 20 labeled items, ten pairs of $(\mathbf{x}, y) = (-1, 0), (1, 1)$
- 22 subjects. Test items drawn from two-component GMM. Two conditions:
  - **L shifted**: GMM $\mu_1 = -1.43, \mu_2 = 0.57$
  - **R shifted**: GMM $\mu_1 = -0.57, \mu_2 = 1.43$





- Early (in first 50 test items) decision boundaries the same
- Late (after 700 test items) boundaries shifted according to condition

Test-Item Effect as Semi-Supervised Learning

Standard human category learning models in psychology cannot explain test-item effects

Test-Item Effect as Semi-Supervised Learning

Standard human category learning models in psychology cannot explain test-item effects

1. exemplar model $\approx$ nonparametric kernel regression

Test-Item Effect as Semi-Supervised Learning

Standard human category learning models in psychology cannot explain test-item effects

1. exemplar model $\approx$ nonparametric kernel regression
2. prototype model $\approx$ Gaussian mixture model

Test-Item Effect as Semi-Supervised Learning

Standard human category learning models in psychology cannot explain test-item effects

1. exemplar model $\approx$ nonparametric kernel regression
2. prototype model $\approx$ Gaussian mixture model
3. rational model of categorization $\approx$ Dirichlet process mixture model

Test-Item Effect as Semi-Supervised Learning

Standard human category learning models in psychology cannot explain test-item effects

1. exemplar model $\approx$ nonparametric kernel regression
2. prototype model $\approx$ Gaussian mixture model
3. rational model of categorization $\approx$ Dirichlet process mixture model

We propose semi-supervised extensions to these models

- incremental (online) learning to better fit human experience
- minimum number of parameters to prevent overfitting

Model 1: Semi-Supervised Exemplar Model

- Extends the generalized context model (Nosofsky, 1986)
- Self-training Nadaraya-Watson kernel estimator

Model 1: Semi-Supervised Exemplar Model

- Extends the generalized context model (Nosofsky, 1986)
- Self-training Nadaraya-Watson kernel estimator

Parameter: kernel bandwidth h
for $n = 1, 2, \ldots$ **do**
    Receive x_n,

Model 1: Semi-Supervised Exemplar Model

- Extends the generalized context model (Nosofsky, 1986)
- Self-training Nadaraya-Watson kernel estimator

Parameter: kernel bandwidth h

for $n = 1, 2, \ldots$ **do**

Receive x_n, predict its label by thresholding

$r(x_n) = \sum_{i=1}^{n-1} \frac{K(\frac{x_n - x_i}{h})}{\sum_{j=1}^{n-1} K(\frac{x_n - x_j}{h})} \hat{y}_i$ at 0.5

Model 1: Semi-Supervised Exemplar Model

- Extends the generalized context model (Nosofsky, 1986)
- Self-training Nadaraya-Watson kernel estimator

Parameter: kernel bandwidth h

for $n = 1, 2, \ldots$ **do**

Receive x_n, predict its label by thresholding

$r(x_n) = \sum_{i=1}^{n-1} \frac{K(\frac{x_n - x_i}{h})}{\sum_{j=1}^{n-1} K(\frac{x_n - x_j}{h})} \hat{y}_i$ at 0.5

Receive y_n (may be unlabeled), update model:

if y_n is unlabeled **then**

$\hat{y}_n = r(x_n)$

else

$\hat{y}_n = y_n$

end if

end for

Model 2: Semi-Supervised Prototype Model

- Extends prototype models (Posner & Keele, 1968)
- Incremental EM on GMM (Neal & Hinton, 1998), but without revisiting old items

Model 2: Semi-Supervised Prototype Model

- Extends prototype models (Posner & Keele, 1968)
- Incremental EM on GMM (Neal & Hinton, 1998), but without revisiting old items
- Track parameters of GMM via sufficient statistics
  - If input (x, y) labeled, its contribution to sufficient statistics is $\tilde{\phi}(x, y) = (1 - y, (1 - y)x, (1 - y)x^2, y, yx, yx^2)$

Model 2: Semi-Supervised Prototype Model

- Extends prototype models (Posner & Keele, 1968)
- Incremental EM on GMM (Neal & Hinton, 1998), but without revisiting old items
- Track parameters of GMM via sufficient statistics
  - If input (x, y) labeled, its contribution to sufficient statistics is $\tilde{\phi}(x, y) = (1 - y, (1 - y)x, (1 - y)x^2, y, yx, yx^2)$
  - If input x unlabeled, it is

$$\mathbb{E}_{y \sim q}[\tilde{\phi}(x, y)] = \sum_{y=0,1} q(y)\tilde{\phi}(x, y)$$

where $q(y) = p(y|x, \theta)$ is the label posterior under the current model

Model 2: Semi-Supervised Prototype Model

- Extends prototype models (Posner & Keele, 1968)
- Incremental EM on GMM (Neal & Hinton, 1998), but without revisiting old items
- Track parameters of GMM via sufficient statistics
  - If input (x, y) labeled, its contribution to sufficient statistics is $\tilde{\phi}(x, y) = (1 - y, (1 - y)x, (1 - y)x^2, y, yx, yx^2)$
  - If input x unlabeled, it is

$$\mathbb{E}_{y \sim q}[\tilde{\phi}(x, y)] = \sum_{y=0,1} q(y)\tilde{\phi}(x, y)$$

    where $q(y) = p(y|x, \theta)$ is the label posterior under the current model
  - Initialize sufficient statistics as $\phi = (n_0, 0, n_0, n_0, 0, n_0)$: n_0 pseudo items with mean 0 and variance 1.
  - n_0 is the only parameter.

Model 3: Semi-Supervised Rational Model of Categorization (RMC)

- Extends RMC (Anderson 1990, Griffiths et al. 2008)
- Dirichlet Process Mixture Model (DPMM) marginalizd over y

Model 3: Semi-Supervised Rational Model of Categorization (RMC)

- Extends RMC (Anderson 1990, Griffiths et al. 2008)
- Dirichlet Process Mixture Model (DPMM) marginalizd over y
  - stack $[x; y]$ and use a single global DPMM (key difference to Aclass (Mansinghka et al. 2007))

Model 3: Semi-Supervised Rational Model of Categorization (RMC)

- Extends RMC (Anderson 1990, Griffiths et al. 2008)
- Dirichlet Process Mixture Model (DPMM) marginalizd over y
  - stack $[x; y]$ and use a single global DPMM (key difference to Aclass (Mansinghka et al. 2007))
  - $G \sim DP(G_0, \alpha_2)$
    - base measure $G_0 =$ Normal-Gamma $\times$ Beta (conjugate priors for Normal and binomial)
    - α_2 is the only parameter

Model 3: Semi-Supervised Rational Model of Categorization (RMC)

- Extends RMC (Anderson 1990, Griffiths et al. 2008)
- Dirichlet Process Mixture Model (DPMM) marginalizd over y
  - stack $[x; y]$ and use a single global DPMM (key difference to Aclass (Mansinghka et al. 2007))
  - $G \sim DP(G_0, \alpha_2)$
    - base measure $G_0 =$ Normal-Gamma $\times$ Beta (conjugate priors for Normal and binomial)
    - α_2 is the only parameter
  - $\theta_1 \ldots \theta_n \sim G$, where $\theta = (\mu, \lambda, p)$
    - μ, λ the mean and precision of a Gaussian for the x component
    - p the "head" probability for the y component

Model 3: Semi-Supervised Rational Model of Categorization (RMC)

- Extends RMC (Anderson 1990, Griffiths et al. 2008)
- Dirichlet Process Mixture Model (DPMM) marginalizd over y
  - stack $[x; y]$ and use a single global DPMM (key difference to Aclass (Mansinghka et al. 2007))
  - $G \sim DP(G_0, \alpha_2)$
    - base measure $G_0 =$ Normal-Gamma $\times$ Beta (conjugate priors for Normal and binomial)
    - α_2 is the only parameter
  - $\theta_1 \ldots \theta_n \sim G$, where $\theta = (\mu, \lambda, p)$
    - μ, λ the mean and precision of a Gaussian for the x component
    - p the "head" probability for the y component
  - $(x_i, y_i) \sim F(x, y|\theta_i)$, $F =$ Gaussian $\times$ Bernoulli

Particle Filtering for Semi-Supervised RMC

Particle Filtering for Semi-Supervised RMC

- Introduce cluster index z

Particle Filtering for Semi-Supervised RMC

- Introduce cluster index z
- Integrate out θ and G via particle filtering

Particle Filtering for Semi-Supervised RMC

- Introduce cluster index z
- Integrate out θ and G via particle filtering
- Each particle is a vector of indices $z_{1:n-1}$

Particle Filtering for Semi-Supervised RMC

- Introduce cluster index z
- Integrate out θ and G via particle filtering
- Each particle is a vector of indices $z_{1:n-1}$
- "Grow" particle by z_n, weight proportional to likelihood

$$P(y_{n-1} \mid z_{1:n-1}, y_{1:n-2}) P(z_n \mid z_{1:n-1}) P(x_n \mid z_n, z_{1:n-1}, x_{1:n-1})$$

Particle Filtering for Semi-Supervised RMC

- Introduce cluster index z
- Integrate out θ and G via particle filtering
- Each particle is a vector of indices $z_{1:n-1}$
- "Grow" particle by z_n, weight proportional to likelihood

$$P(y_{n-1} \mid z_{1:n-1}, y_{1:n-2})P(z_n \mid z_{1:n-1})P(x_n \mid z_n, z_{1:n-1}, x_{1:n-1})$$

- For semi-supervised DPMM, the y term is a beta-binomial with marginalization

$$P(y_{n-1} \mid z_{1:n-1}, y_{1:n-2}) = \frac{c_1 + \alpha_1}{c_0 + c_1 + \alpha_1 + \beta_1}$$

  - If y_{n-1} unlabeled, define the probability to be 1

Particle Filtering for Semi-Supervised RMC

- Introduce cluster index z
- Integrate out θ and G via particle filtering
- Each particle is a vector of indices $z_{1:n-1}$
- "Grow" particle by z_n, weight proportional to likelihood

$$P(y_{n-1} \mid z_{1:n-1}, y_{1:n-2}) P(z_n \mid z_{1:n-1}) P(x_n \mid z_n, z_{1:n-1}, x_{1:n-1})$$

- For semi-supervised DPMM, the y term is a beta-binomial with marginalization

$$P(y_{n-1} \mid z_{1:n-1}, y_{1:n-2}) = \frac{c_1 + \alpha_1}{c_0 + c_1 + \alpha_1 + \beta_1}$$

  - If y_{n-1} unlabeled, define the probability to be 1
  - If some of $y_{1:n-2}$ unlabeled, skip them in counting

$$c_1 = \sum_{i=1}^{n-2} \delta(z_i, z_{n-1}) \delta(y_i, 1) \quad c_0 = \sum_{i=1}^{n-2} \delta(z_i, z_{n-1}) \delta(y_i, 0)$$

Parameter Tuning for All Three Models

- Divide subjects into training and test groups
- Maximize training group human prediction likelihood:

$$\theta^* = \arg\max_\theta \ell_{tr}(\theta) \equiv \sum_{s \in tr} \sum_n \log P(f(\mathbf{x}_n)^{[s]} \mid x_{1:n}^{[s]}, y_{1:n-1}^{[s]}, \theta)$$

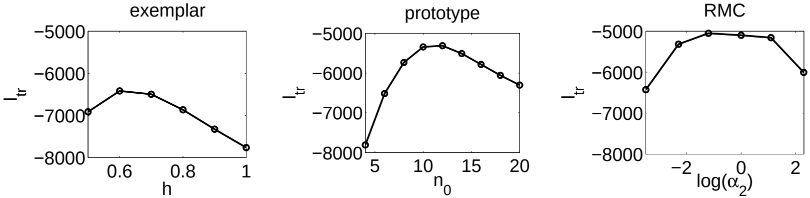
where θ is h, n_0, α_2 for the three models, respectively.

Parameter Tuning for All Three Models

- Divide subjects into training and test groups
- Maximize training group human prediction likelihood:

$$\theta^* = \arg\max_{\theta} \ell_{tr}(\theta) \equiv \sum_{s \in tr} \sum_{n} \log P(f(\mathbf{x}_n)^{[s]} \mid x_{1:n}^{[s]}, y_{1:n-1}^{[s]}, \theta)$$

where θ is h, n_0, α_2 for the three models, respectively.



Model Fitting Results

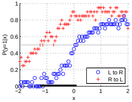
Performance comparison on **test group**:

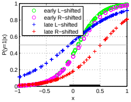
	SSL exemplar	SSL prototype	SSL RMC
θ^*	$h = 0.6$	$n_0 = 12$	$\alpha_2 = 0.3$
$\ell_{te}(\theta^*)$	-3727	-2460	**-2169**

Semi-supervised RMC has the best fit, semi-supervised exemplar model the worst.

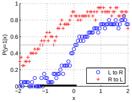
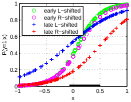
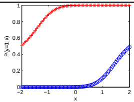
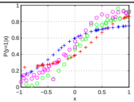
Model Predictions

human



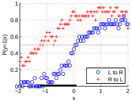


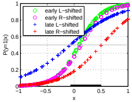
Model Predictions

human		
SSL exemplar	 $h^* = 0.6$	 $h^* = 0.6$

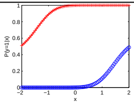
Model Predictions

human

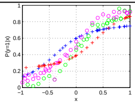




SSL exemplar

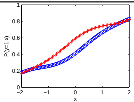


$h^* = 0.6$

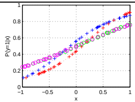


$h^* = 0.6$

SSL prototype



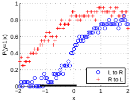
$n_0^* = 12$

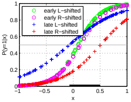


$n_0^* = 12$

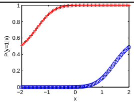
Model Predictions

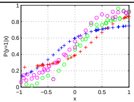
human





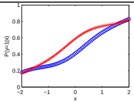
SSL exemplar

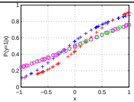




$h^* = 0.6$ $h^* = 0.6$

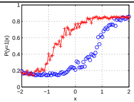
SSL prototype

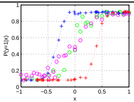




$n_0^* = 12$ $n_0^* = 12$

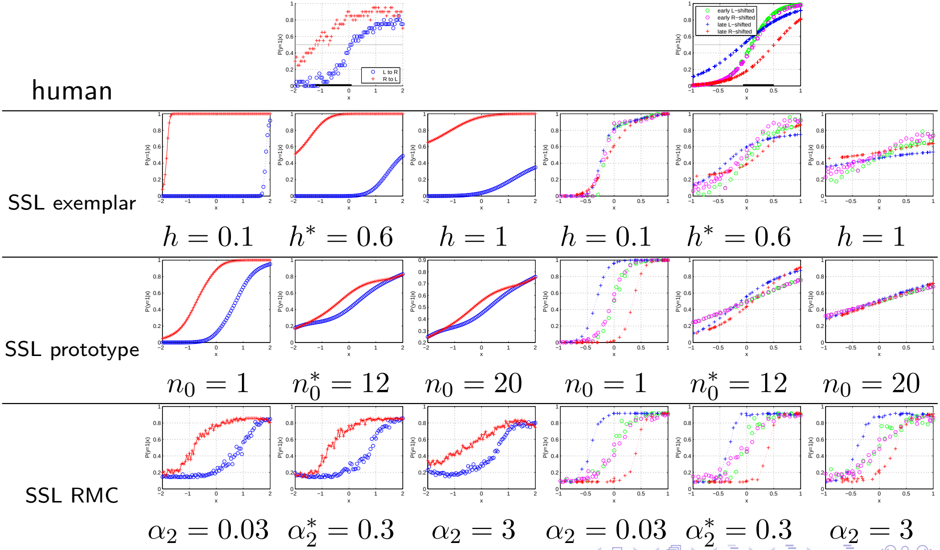
SSL RMC





$\alpha_2^* = 0.3$ $\alpha_2^* = 0.3$

Model Predictions



Attempts to Save Semi-Supervised Exemplar Model

- What if we down-weight unlabeled items?

$$r(x) = \sum_{i=1}^{n} \frac{w_i K(\frac{x-x_i}{h})}{\sum_{j=1}^{n} w_i K(\frac{x-x_j}{h})} y_i$$

$w_i = 1$ if $\mathbf{x}_i$ labeled, $w_i = w$ otherwise

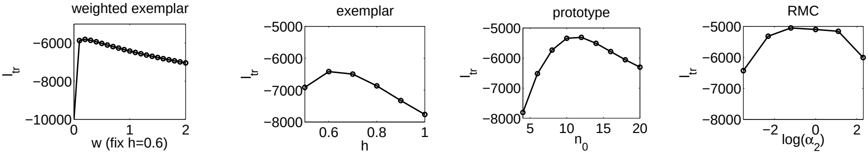
Attempts to Save Semi-Supervised Exemplar Model

- What if we down-weight unlabeled items?

$$r(x) = \sum_{i=1}^{n} \frac{w_i K(\frac{x-x_i}{h})}{\sum_{j=1}^{n} w_i K(\frac{x-x_j}{h})} y_i$$

$w_i = 1$ if $\mathbf{x}_i$ labeled, $w_i = w$ otherwise

- Learned $w = 0.2$. Test group loglik -2934, still worse.



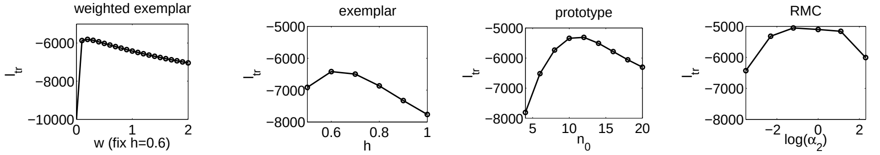
Attempts to Save Semi-Supervised Exemplar Model

- What if we down-weight unlabeled items?

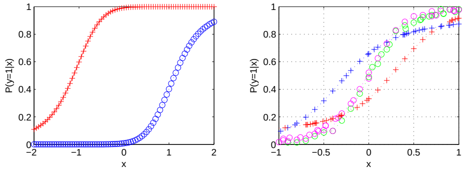
$$r(x) = \sum_{i=1}^{n} \frac{w_i K(\frac{x-x_i}{h})}{\sum_{j=1}^{n} w_i K(\frac{x-x_j}{h})} y_i$$

$w_i = 1$ if $\mathbf{x}_i$ labeled, $w_i = w$ otherwise

- Learned $w = 0.2$. Test group loglik -2934, still worse.



- Model predictions still qualitatively poor:



Conclusions

Contributions

- Test-item effects in humans
- Semi-supervised extension of exemplar, prototype, and ration model of categorization
  - All three models exhibit test-item effects
  - Semi-supervised RMC the best

Conclusions

Contributions

- Test-item effects in humans
- Semi-supervised extension of exemplar, prototype, and ration model of categorization
  - All three models exhibit test-item effects
  - Semi-supervised RMC the best
- Take home message: cognitive psychology ideal application for machine learning.
  - Coming soon: Cognitive Modeling Repository `http://www.cmr.osu.edu/`

This work is supported in part by AFOSR FA9550-09-1-0313, NSF IIS-0916038, IIS-0953219, DLS/DRM-0745423, and the Wisconsin Alumni Research Foundation.