Humans Perform Semi-Supervised Classification Too

Xiaojin Zhu Timothy Rogers*
Ruichen Qian Chuck Kalish*

Department of Computer Sciences
Department of Psychology*
University of Wisconsin, Madison, USA

AAAI 2007

A camping story

A camping story

A camping story

badger

A camping story

A camping story

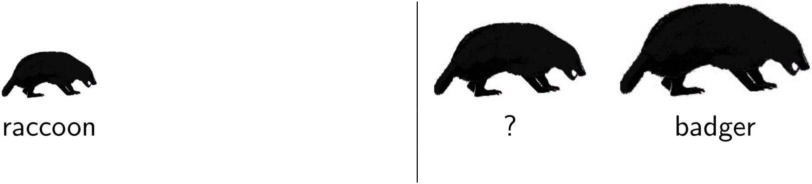
A camping story

raccoon

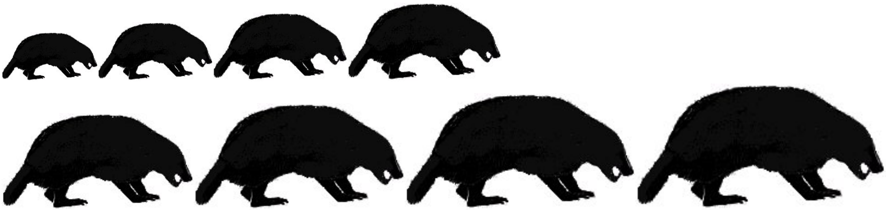
A camping story

?

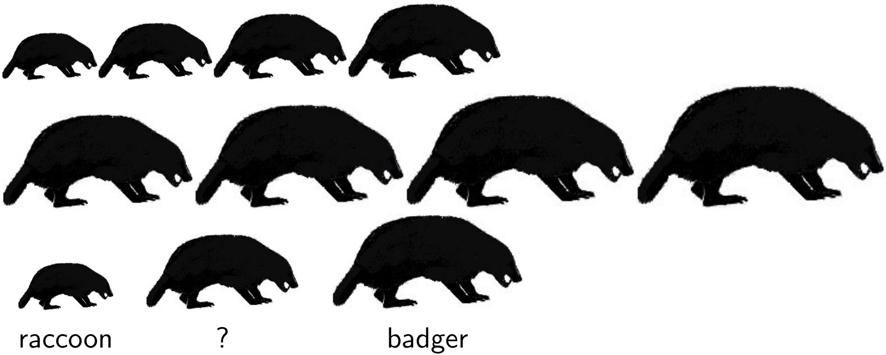
A camping story



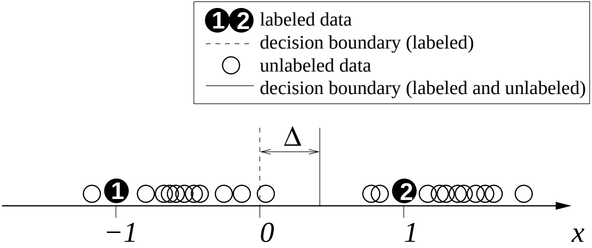
A camping story



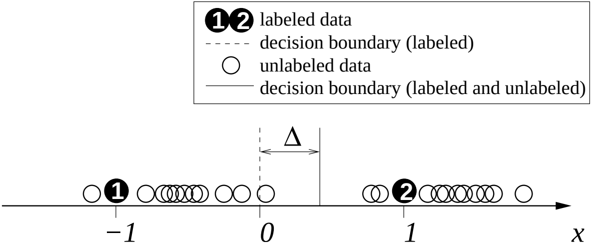
A camping story



The semi-supervised learning task



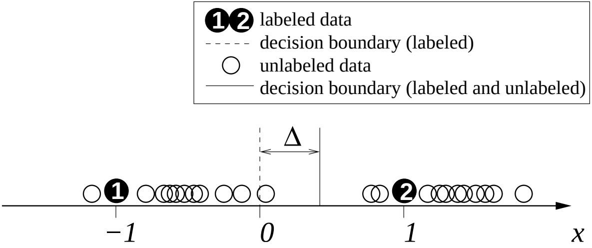
The semi-supervised learning task



- Assuming each class is a coherent group (e.g. Gaussian), semi-supervised machine learning predicts decision boundary shift.

[Castelli & Cover 96; Ratsaby & Venkatesh 95; Nigam et al. 00]

The semi-supervised learning task



- Assuming each class is a coherent group (e.g. Gaussian), semi-supervised machine learning predicts decision boundary shift.
  [Castelli & Cover 96; Ratsaby & Venkatesh 95; Nigam et al. 00]
- Do we humans shift decision boundary too?

Human learning: a behavioral experiment

Goal

Determine human decision boundaries for:

- labeled data only vs. labeled and unlabeled data
- same labeled data, different unlabeled data

Participants and materials

- 22 University of Wisconsin students

Participants and materials

- 22 University of Wisconsin students
- Novel stimuli displayed one at a time (nothing stays on screen)

Participants and materials

- 22 University of Wisconsin students
- Novel stimuli displayed one at a time (nothing stays on screen)
- Stimuli parameterized by a single parameter

Participants and materials

- 22 University of Wisconsin students
- Novel stimuli displayed one at a time (nothing stays on screen)
- Stimuli parameterized by a single parameter
- Told stimuli are microscopic pollens

Participants and materials

- 22 University of Wisconsin students
- Novel stimuli displayed one at a time (nothing stays on screen)
- Stimuli parameterized by a single parameter
- Told stimuli are microscopic pollens
- Press B or N to classify

Participants and materials

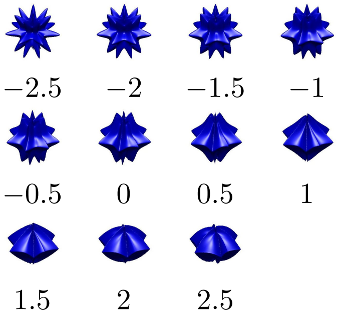
- 22 University of Wisconsin students
- Novel stimuli displayed one at a time (nothing stays on screen)
- Stimuli parameterized by a single parameter
- Told stimuli are microscopic pollens
- Press B or N to classify
- Label: audio feedback

Participants and materials

- 22 University of Wisconsin students
- Novel stimuli displayed one at a time (nothing stays on screen)
- Stimuli parameterized by a single parameter
- Told stimuli are microscopic pollens
- Press B or N to classify
- Label: audio feedback
- No audio feedback for unlabeled data

Visual stimuli

Stimuli parameterized by a continuous variable x.



Experiment procedure

- Half L-subjects, half R-subjects

Experiment procedure

- Half L-subjects, half R-subjects
- Each subject sees 6 blocks of stimuli

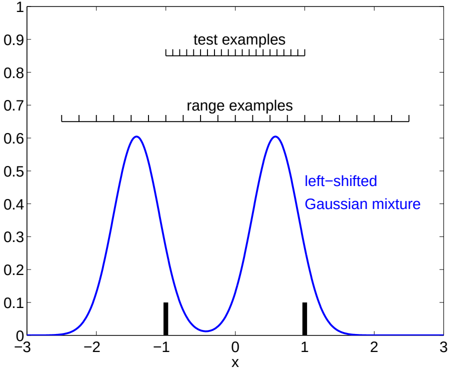
Experiment procedure

- Half L-subjects, half R-subjects
- Each subject sees 6 blocks of stimuli
- Order within each block randomized

Experiment procedure

- Half L-subjects, half R-subjects
- Each subject sees 6 blocks of stimuli
- Order within each block randomized
- Record their decisions and response times

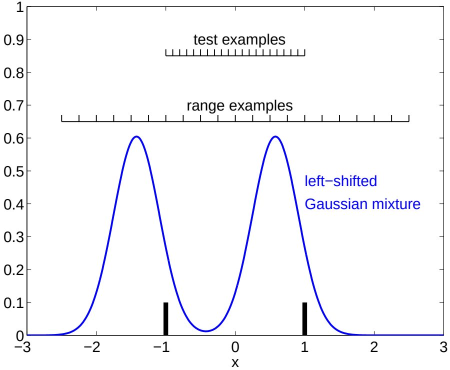
The 6 blocks of stimuli



stimuli

1. (labeled) 10 $(x = 1, B)$, 10 $(x = -1, N)$. The only labeled block.

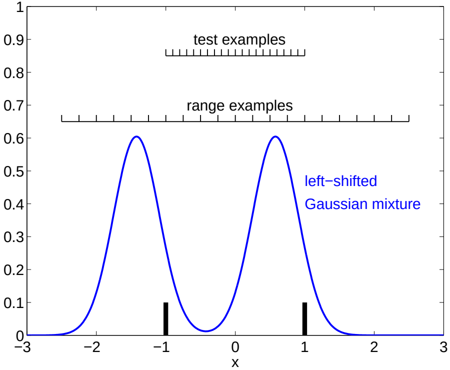
The 6 blocks of stimuli



stimuli

1. (labeled) 10 $(x = 1, B)$, 10 $(x = -1, N)$. The only labeled block.
2. (test-1) $x = -1, -0.9, \ldots, 0.9, 1$

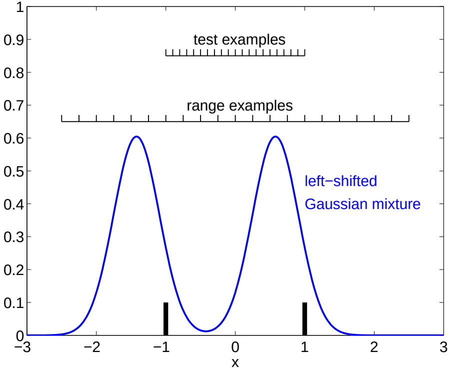
The 6 blocks of stimuli



stimuli

1. (labeled) 10 $(x = 1, B)$, 10 $(x = -1, N)$. The only labeled block.
2. (test-1) $x = -1, -0.9, \ldots, 0.9, 1$
3. (unlabeled-1) 230 stimuli $\sim$ offset 2 Gaussian, left- or right-shifted. 21 range stimuli evenly in $[-2.5, 2.5]$

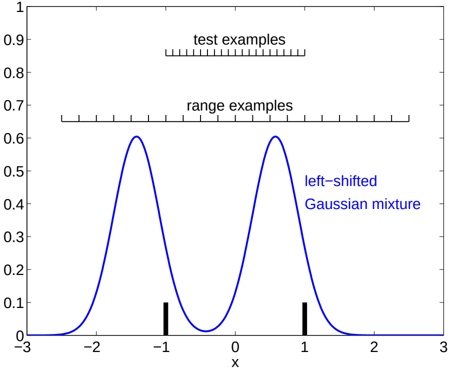
The 6 blocks of stimuli



stimuli

1. (labeled) 10 $(x = 1, B)$, 10 $(x = -1, N)$. The only labeled block.
2. (test-1) $x = -1, -0.9, \ldots, 0.9, 1$
3. (unlabeled-1) 230 stimuli $\sim$ offset 2 Gaussian, left- or right-shifted. 21 range stimuli evenly in $[-2.5, 2.5]$
4. (unlabeled-2) similar to block 3
5. (unlabeled-3) similar to block 3

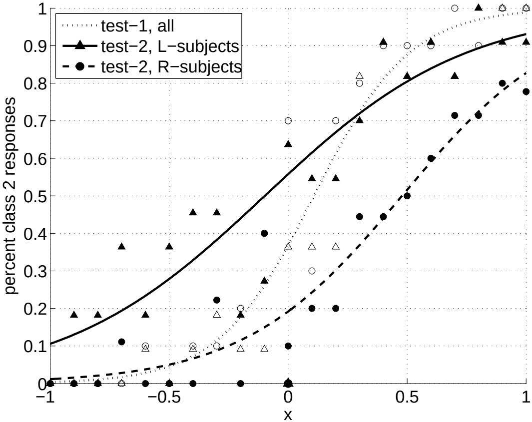
The 6 blocks of stimuli



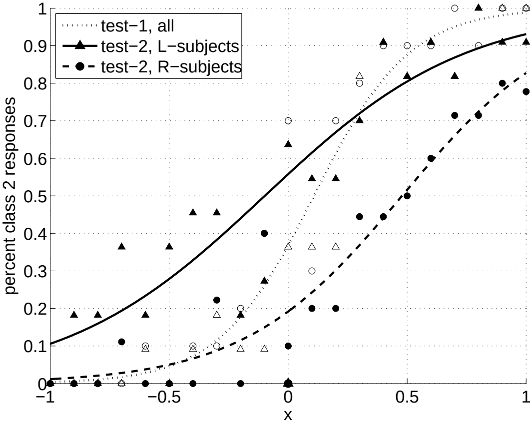
stimuli

1. (labeled) 10 $(x = 1, B)$, 10 $(x = -1, N)$. The only labeled block.
2. (test-1) $x = -1, -0.9, \ldots, 0.9, 1$
3. (unlabeled-1) 230 stimuli $\sim$ offset 2 Gaussian, left- or right-shifted. 21 range stimuli evenly in $[-2.5, 2.5]$
4. (unlabeled-2) similar to block 3
5. (unlabeled-3) similar to block 3
6. (test-2) $x = -1, -0.9, \ldots, 0.9, 1$

Observation 1: Unlabeled data affects decision boundary



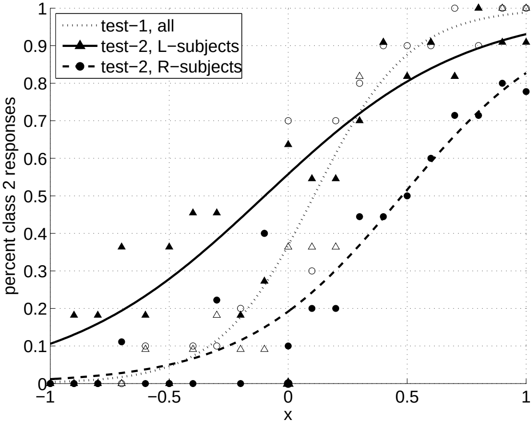
Observation 1: Unlabeled data affects decision boundary



Decision boundary:

- after labeled data (test-1): $x = 0.11$

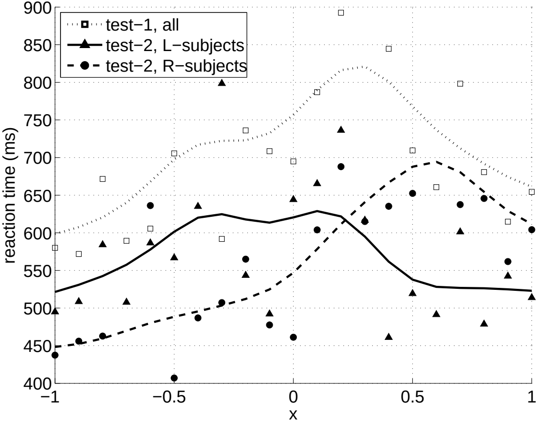
Observation 1: Unlabeled data affects decision boundary



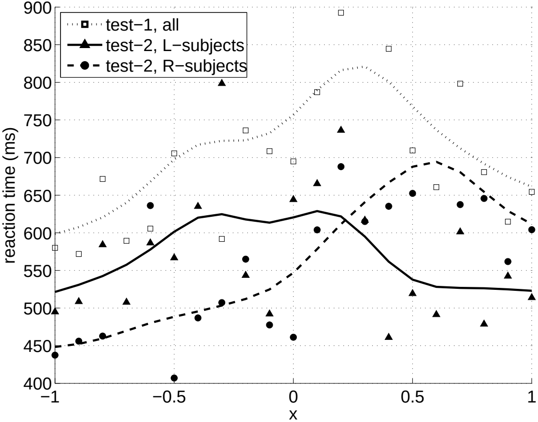
Decision boundary:

- after labeled data (test-1): $x = 0.11$
- after labeled and unlabeled data (test-2):
  L-subjects $x = -0.10$, R-subjects $x = 0.48$

Observation 2: Reaction time reflects boundary shift



Observation 2: Reaction time reflects boundary shift



- Longer reaction time → closer to decision boundary
- Test-2 overall faster, familiarity with experiment
- L-, R-reaction time further support decision boundary shift

Machine learning model

We can explain the human experiment with a 2-component Gaussian Mixture Model (GMM).

The GMM:

$$w_1 N(\mu_1, \sigma_1^2) + w_2 N(\mu_2, \sigma_2^2) \quad , w_1 + w_2 = 1, w_i \geq 0$$

Machine learning model

We can explain the human experiment with a 2-component Gaussian Mixture Model (GMM).

The GMM:

$$w_1 N(\mu_1, \sigma_1^2) + w_2 N(\mu_2, \sigma_2^2) \quad , w_1 + w_2 = 1, w_i \geq 0$$

Prior on parameters θ:

$$w_k \sim \mathrm{Uniform}[0,1], \mu_k \sim \mathrm{N}(0, \infty), \sigma_k^2 \sim \mathrm{Inv}{-}\chi^2(\nu, s^2), k = 1, 2$$

Machine learning model

We can explain the human experiment with a 2-component Gaussian Mixture Model (GMM).

The GMM:

$$w_1 N(\mu_1, \sigma_1^2) + w_2 N(\mu_2, \sigma_2^2) \quad , w_1 + w_2 = 1, w_i \geq 0$$

Prior on parameters θ:

$$w_k \sim \mathrm{Uniform}[0,1], \mu_k \sim \mathrm{N}(0, \infty), \sigma_k^2 \sim \mathrm{Inv}{-}\chi^2(\nu, s^2), k = 1, 2$$

We fit the GMM with the Expectation-Maximization (EM) algorithm on blocks 1,2 vs. on all blocks.

EM

Maximize the objective ($\lambda \leq 1$ weight on unlabeled example)

$$\log p(\theta) + \sum_{i=1}^{l} \log p(x_i, y_i|\theta) + \lambda \sum_{i=l+1}^{n} \log p(x_i|\theta)$$

E-step

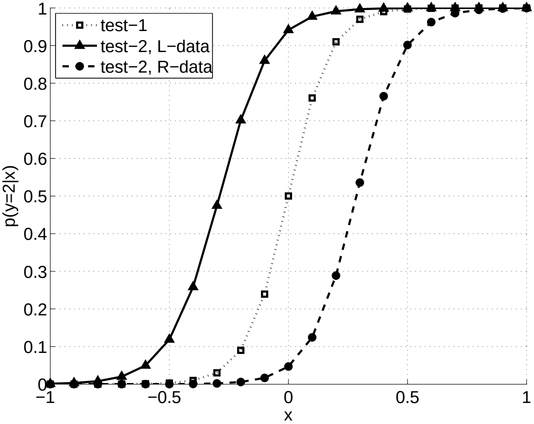
$$q_i(k) \propto w_k \mathrm{N}(x_i; \mu_k, \sigma_k^2), \quad i = l+1, \ldots, n; k = 1, 2$$

M-step

$$\begin{aligned}
\mu_k &= \frac{\sum_{i=1}^{l} \delta(y_i, k) x_i + \lambda \sum_{i=l+1}^{n} q_i(k) x_i}{\sum_{i=1}^{l} \delta(y_i, k) + \lambda \sum_{i=l+1}^{n} q_i(k)} \\
\sigma_k^2 &= \frac{\nu s^2 + \sum_{i=1}^{l} \delta(y_i, k) e_{ik} + \lambda \sum_{i=l+1}^{n} q_i(k) e_{ik}}{\nu + 2 + \sum_{i=1}^{l} \delta(y_i, k) + \lambda \sum_{i=l+1}^{n} q_i(k)} \\
w_k &= \frac{\sum_{i=1}^{l} \delta(y_i, k) + \lambda \sum_{i=l+1}^{n} q_i(k)}{l + \lambda(n - l)}
\end{aligned}$$

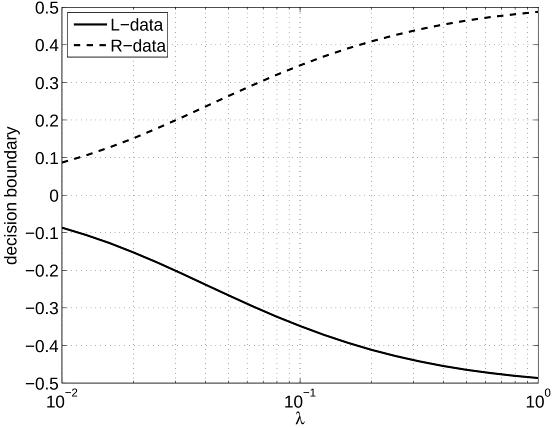
Model fitting result 1

The GMM predicts the decision boundary shift:



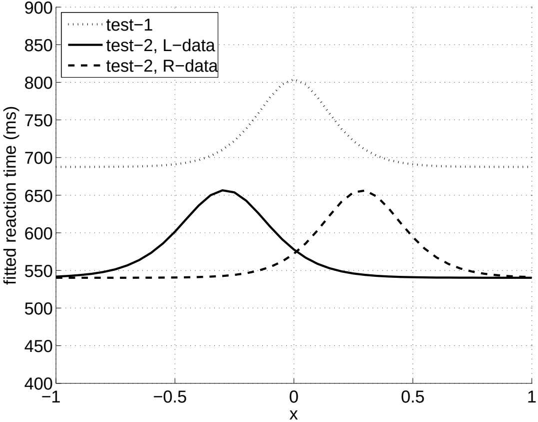
Model fitting result 2

Unlabeled data seem to worth less than labeled data $(\lambda = 0.06)$



Model fitting result 3

The GMM also explains reaction time:



$t = aH(x) + b$, $H(x)$ the entropy of label prediction

Conclusions

- Humans and machines both perform semi-supervised learning.

Conclusions

- Humans and machines both perform semi-supervised learning.
- Flatness on [test-2] not well explained.

Conclusions

- Humans and machines both perform semi-supervised learning.
- Flatness on [test-2] not well explained.
- Co-training, manifold regularization, S3VMs, etc. in humans should be explored.

Conclusions

- Humans and machines both perform semi-supervised learning.
- Flatness on [test-2] not well explained.
- Co-training, manifold regularization, S3VMs, etc. in humans should be explored.
- Further study may lead to new learning algorithms.