AClass: A Simple, Online Probabilistic Classifier

Vikash K. Mansinghka
Computational Cognitive Science Group
MIT BCS/CSAIL
AClass: A Simple, Online Probabilistic Classifier

or

How I learned to stop worrying and love generative models and nonparametric Bayes

Vikash K. Mansinghka
Computational Cognitive Science Group
MIT BCS/CSAIL

Joint work with Daniel Roy, Ryan Rifkin, Joshua Tenenbaum
Outline

- The problem of classification
- Some history and parallels (cogsci and AI)
- AClass: Model and online inference
- Applications to scene analysis
- Using classical AI search ideas for probabilistic inference
Rhetorical Objectives

• Intuitive inductive biases can yield effective classifiers:
  – Prototype finding via nonparametric Bayesian density estimation
  – Tractable online, bounded memory inference via particle filtering

• Classification and its generalizations are:
  – Very important practically
  – (Now, only somewhat) important conceptually/cognitively

• Machine learning and cognitive science have helped each other before, and should again
A Machine Learning Results Teaser

- FIXME: Find good summary figures with our lines higher up than their lines; also maybe a good clustering result
You must be at least this tall to ride... (not really)

- Naïve Bayesian classifier
- Mixture Model or Density estimation
- Prototype or Exemplar models
- Backpropagation
- Kernel Methods
  - Mercer's Theorem
- Condensed Nearest Neighbor
- Chinese Restaurant Process
- AClass
Outline

● The problem of classification
● Some history and parallels (cogsci and AI)
● AClass: Model and online inference
● Applications to scene analysis
● Using classical AI search ideas for probabilistic inference
The problem of classification

- FIXME: Something intuitive; maybe apples vs plums here?
### Machine Learning

**Useful microcosm**  
(spawned semi, active, unsupervised)

Reductions make it somewhat universally applicable

**Lots of approaches do about the same despite decades of work**

**There's a big world out there (structure; direct generation of signals; freedom from the goose chase for “invariant features”)**

### Cognitive Science

**Simplest kinds of concepts**

Classically important; lots of empirical and theoretical work with links to machine learning;  
Clearly highlights importance of induction (and utility of statistics)

**Stresses key representational assumptions**  
(e.g. deterministic vs nondeterministic latent state - essentialism?)

“Labels” are very important and have structure  
(overlapping? cross-cutting? hierarchical?)

**There's a big world out there (think theories)**
The problem of classification

- FIXME: Synthetic distribution figure here (size vs reflectance)
The problem of classification

- FIXME: Decision boundaries vs exploratory data analysis
Outline

- The problem of classification
- Some history and parallels (cogsci and AI)
- AClass: Model and online inference
- Applications to scene analysis
- Using classical AI search ideas for probabilistic inference
<table>
<thead>
<tr>
<th>Machine Learning</th>
<th>Cognitive Science</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GENERATIVE</strong></td>
<td></td>
</tr>
<tr>
<td>Naïve Bayes</td>
<td>Prototype (1 per class)</td>
</tr>
<tr>
<td>Class-conditional Mixture Models</td>
<td>Prototype ((k_c) per class)</td>
</tr>
<tr>
<td>Class-conditional Kernel Density Estimation</td>
<td>Exemplar</td>
</tr>
<tr>
<td><strong>DISCRIMINATIVE</strong></td>
<td></td>
</tr>
<tr>
<td>Perceptron</td>
<td>PDP/Connectionism</td>
</tr>
<tr>
<td>Logistic Regression/Neural Networks</td>
<td>“similarity”-based generalization</td>
</tr>
<tr>
<td>Kernel Methods (e.g. SVMs) and</td>
<td></td>
</tr>
<tr>
<td>nearest neighbor</td>
<td>“rule”-based generalization</td>
</tr>
<tr>
<td>Inductive Logic Programming</td>
<td></td>
</tr>
<tr>
<td>Machine Learning</td>
<td>Associations</td>
</tr>
<tr>
<td>------------------</td>
<td>--------------</td>
</tr>
<tr>
<td><strong>GENERATIVE</strong></td>
<td>Simple, fast; “low” asymptotic accuracy</td>
</tr>
<tr>
<td>Naïve Bayes</td>
<td>Cool; Hard to train (nested EM; crossvalidation)</td>
</tr>
<tr>
<td>Class-conditional Mixture Models</td>
<td>Often generalizes poorly; hard to set bandwidth</td>
</tr>
<tr>
<td>Class-conditional Kernel Density Estimation</td>
<td></td>
</tr>
<tr>
<td><strong>DISCRIMINATIVE</strong></td>
<td>Simple, very fast; pretty bad</td>
</tr>
<tr>
<td>Perceptron</td>
<td>Black art; often good empirically</td>
</tr>
<tr>
<td>Logistic Regression/Neural Networks</td>
<td>Black art; nice math; hard for very big datasets</td>
</tr>
<tr>
<td>Kernel Methods (e.g. SVMs) and nearest neighbor</td>
<td>Weird</td>
</tr>
<tr>
<td>Inductive Logic Programming</td>
<td></td>
</tr>
</tbody>
</table>

Black art; nice math; hard for very big datasets

Weird
Outline

• The problem of classification
• Some history and parallels (cogsci and AI)
• AClass: Model and online inference
• Applications to scene analysis
• Using classical AI search ideas for probabilistic inference
AClass – The Model

- FIXME: Graphical model on board
Comparison to Prior Art: “Bigoted Bayes”

- FIXME: Comparison w/ naïve Bayes
Details: Chinese Restaurant Processes

- FIXME: CRP picture, generative process; CRP mixtures
Details: Particle filters for CRPs

- FIXMEFIXME: Fearnhead figure of results; particle filter pseudocode
AClass – The Algorithm

```
AClass-Train(\(O\): outer-filter, \(I\): class-label, \(y\): observation)
1  foreach particle \(p\) in \(O\)
2    \(w[p]\) ← \text{PREDICTIVE-DENSITY}(p, I[I], y)
3  \text{Train-CRF}(p, I[I], y)
4  \text{Resample-Particles}(O, w)

AClass-Train-Unlabelled(\(y\): observation)
1  foreach particle \(p\) in \(O\)
2    foreach label \(l\) in 1, 2, \ldots, \(L\)
3      \(\text{prob}[l] = \frac{m[l]+\gamma}{\sum (m[l]+L\cdot \gamma)}\)
4    \text{PREDICTIVE-DENSITY}(p, I[I], y)
5    \(w[p] ← \sum(\text{prob})\)
6    \(l ← \text{Sample-Discrete}(\text{prob})\)
7    \(m[l] ← m[l]+1\)
8  \text{Train-CRF}(p, I[I], y)
9  \text{Resample-Particles}(O, w)

AClass-Test(\(O\): outer-filter, \(y\): observation)
1  foreach particle \(p\) in \(O\)
2    foreach label \(l\) in 1, 2, \ldots, \(L\)
3      \(\text{prob}[l][l] = \frac{m[l]+\gamma}{\sum (m[l]+L\cdot \gamma)}\)
4    \text{PREDICTIVE-DENSITY}(p, I[I], y)
5    \(\text{prob}[p]\) ← \text{Normalise}(\text{prob}[p][l])
6  return \text{Average}_{p}(\text{prob}[p][l])
```

```
Train-CRF(\(I\): inner-filter, \(y\): observations)
1  foreach particle \(p\) in \(I\)
2    foreach group \(g\) in 1, 2, \ldots, \(p\).numGroups
3      \(\text{prob}[g] = \frac{n[g]}{\sum(n[g]+a)}\) · \text{POSTERIOR-PREDICTIVE}(p. state\([g]\), y)
4    \(\text{prob}[g+1] = \frac{n[g]+\gamma}{\sum(n[g]+\gamma)}\) · \text{POSTERIOR-PREDICTIVE}(p. state\([g+1]\), y)
5    \(w[p] ← \sum_{a}(\text{prob})\)
6    \(g' ← \text{Sample-Discrete}(\text{prob})\)
7    \(n[g'] ← n[g'] + 1\)
8    \text{Update-Sufficient-Statistics}(p, g', y)
9  \text{Resample-Particles}(I, w)

PREDICTIVE-DENSITY(\(I\): inner-filter, \(y\): observation)
1  foreach particle \(p\) in \(I\)
2    foreach group \(g\) in 1, 2, \ldots, \(p\).numGroups
3      \(\text{prob}[p][g] = \frac{n[g]}{\sum(n[g]+a)}\) · \text{POSTERIOR-PREDICTIVE}(p. state\([g]\), y)
4    \(\text{prob}[p][g+1] = \frac{n[g]+\gamma}{\sum(n[g]+\gamma)}\) · \text{POSTERIOR-PREDICTIVE}(p. state\([g+1]\), y)
5    \(\text{prob}[p] ← \sum_{a}(\text{prob}[p][g])\)
6  return \text{Average}_{p}(\text{prob}[p])
```
Nice Properties of AClass

• Model:
  – Flexibly interpolates between finding prototypes and exemplars as data dictates
  – Nonlinear decision boundary derived from intuitive inductive bias

• Algorithm:
  – Online (and bounded memory)
  – Parallelizable
  – Simple, short
Some Preliminary Results

- FIXME: Results figure
Outline

• The problem of classification
• Some history and parallels (cogsci and AI)
• AClass: Model and online inference
• Applications to scene analysis
• Using classical AI search ideas for probabilistic inference
Applications to Scene Analysis

• FIXME: Have figures here; talk for now
AI Search and Probabilistic Inference: The New Frankenstein

- FIXME: Discuss sequentialization approximation; how AClass is a special case, etc.
Conclusions

• Intuitive inductive biases can yield effective classifiers:
  – Prototype finding via nonparametric Bayesian density estimation
  – Tractable online, bounded memory inference via particle filtering

• Classification and its generalizations are:
  – Very important practically
  – (Now, only somewhat) important conceptually/cognitively

• Machine learning and cognitive science have helped each other before, and should again