Week 1, Video 5

Case Study – San Pedro
Case Study of Classification

- With educational data
- Thousands of examples to choose from
- This example is one I know particularly well
Case Study of Classification

Research Goal

- Can we predict student college attendance
- Based on student engagement and learning in middle school mathematics
- Using fine-grained indicators distilled from interactions with educational software in middle school (~5 years earlier)
Why?

- We can infer engagement and learning in middle school, which supports
  Automated intervention
  Providing actionable info to teachers and school leaders

- But which indicators of engagement and learning really matter?
  Can we find indicators that a student is at-risk, that we can act on, before problem becomes critical?
ASSISTments
Log Data

- 3,747 students
  - In 3 school districts in Massachusetts
    - 1 urban
    - 2 suburban
- Completed 494,150 math problems
  - Working approximately 1 class period a week for the entire year
- Making 2,107,108 problem-solving attempts or hint requests in ASSISTments
- Between 2004-2007
Data set

- Records about whether student eventually attended college
- 58% of students in sample attended college
Automated Detectors

- A number of automated detectors were applied to the data from ASSISTments.
- These detectors had themselves been previously developed using prediction modeling and were published in previous papers, including (Pardos et al., 2013).
- Building a detector and then using it in another analysis is called *discovery with models*.
Automated Detectors

- Learning
  Bayesian Knowledge Tracing; we’ll discuss this later in the course
Disengagement Detectors (No sensors! Just log files!)

- **Gaming the System**
  - Intentional misuse of educational software
  - Systematic Guessing or Rapid Hint Requests

- **Off-Task Behavior**
  - Stopping work in educational software to do unrelated task
  - Does *not* include talking to the teacher or another student about math; these can be distinguished by behavior before and after a pause

- **Carelessness**
  - Making errors despite knowing skill
Affect Detectors (No sensors! Just log files!)

- Boredom
- Frustration
- Confusion
- Engaged Concentration
College Attendance Model

- Predict whether a student attended college from a student’s year-long average according to the detectors
  - **Logistic Regression** Classifier (binary data)
  - Cross-validated at the student-level

  We’ll discuss this next week
## Individual Feature Predictiveness

<table>
<thead>
<tr>
<th>Feature</th>
<th>College</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student Knowledge</td>
<td>NO</td>
<td>0.292</td>
<td>0.151</td>
<td>-15.481</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>YES</td>
<td>0.378</td>
<td>0.180</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correctness</td>
<td>NO</td>
<td>0.382</td>
<td>0.161</td>
<td>-17.793</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>YES</td>
<td>0.483</td>
<td>0.182</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boredom</td>
<td>NO</td>
<td>0.287</td>
<td>0.045</td>
<td>5.974</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>YES</td>
<td>0.278</td>
<td>0.047</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Engaged Concentration</td>
<td>NO</td>
<td>0.483</td>
<td>0.041</td>
<td>-11.979</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>YES</td>
<td>0.500</td>
<td>0.044</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Confusion</td>
<td>NO</td>
<td>0.130</td>
<td>0.054</td>
<td>5.686</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>YES</td>
<td>0.120</td>
<td>0.052</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Individual Feature Predictiveness

<table>
<thead>
<tr>
<th>Feature</th>
<th>College</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Off-Task</td>
<td>NO</td>
<td>0.304</td>
<td>0.119</td>
<td>1.184</td>
<td>0.237</td>
</tr>
<tr>
<td></td>
<td>YES</td>
<td>0.300</td>
<td>0.116</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gaming</td>
<td>NO</td>
<td>0.041</td>
<td>0.062</td>
<td>8.862</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>YES</td>
<td>0.026</td>
<td>0.044</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carelessness</td>
<td>NO</td>
<td>0.132</td>
<td>0.066</td>
<td>-13.361</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>YES</td>
<td>0.165</td>
<td>0.077</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of First Actions</td>
<td>NO</td>
<td>114.50</td>
<td>91.771</td>
<td>-8.673</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>(Proxy for Attendance)</td>
<td>YES</td>
<td>144.56</td>
<td>113.357</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Full Model

- $A' = 0.686$, Kappa $= 0.247$
- $\chi^2 (df = 6, N = 3747) = 386.502$, $p < 0.001$ (computed for a non-cross-validated model)
- $R^2$ (Cox & Snell) $= 0.098$, $R^2$ (Nagelkerke) $= 0.132$
- Overall accuracy $= 64.6\%$; Precision $= 66.4\%$; Recall rate $= 78.3\%$
CollegeEnrollment =
+ 1.119 StudentKnowledge
+ 0.698 Correctness
+ 0.261 NumFirstActions
– 1.145 Carelessnessness
+ 0.217 Confusion
+ 0.169 Boredom
+ 0.351
Flipped Signs

CollegeEnrollment =
+ 1.119 StudentKnowledge
+ 0.698 Correctness
+ 0.261 NumFirstActions
– 1.145 Carelessnessness
+ 0.217 Confusion
+ 0.169 Boredom
+ 0.351
Implications

- Carelessness is bad… once we take knowledge into account
- Boredom is not a major problem… among knowledgeable students
  
  When unsuccessful bored students are removed, all that may remain are those who become bored because material may be too easy
  
  Does not mean boredom is a good thing!
Gaming the System drops out of model

Probably because gaming substantially hurts learning

But just because Gaming->Dropout is likely mediated by learning, doesn’t mean gaming doesn’t matter!

- 0.34 $\sigma$ effect
Implications

- Off-Task Behavior is not such a big deal
  How much effort goes into stopping it?
  Past meta-analyses find small significant effect on short-term measures of learning
  - But not when collaborative learning is occurring?
Implications

- In-the-moment interventions provided by software (or suggested by software to teachers) may have unexpectedly large effects, if they address boredom, confusion, carelessness, gaming the system
Next Lecture

- Less conservative classification algorithms