Model-Based Machine Learning Tutorial with Infer.NET

John Guiver
Model-based machine learning

- Data scientist specifies a generative model (e.g. simulator)
- ML algorithm is generated automatically
- Ideally:
  - Writing simulator only requires knowledge of domain (not ML)
  - Generated algorithms are competitive with traditional algorithms (speed and accuracy)
Roadmap

- Understand and explore a specific data set
- Iteratively:
  - Build up a model of the data
  - Understand the deficiencies of the model by analysing the results
  - Come up with ideas for how to improve the model by revisiting the data
Model-based Machine Learning

A model
⇔
A set of assumptions about how the data was generated
⇔
A simulation of that generation process

Explore different models ⇔ Explore different assumptions
The modelling workflow
**Graphical Model**

- A factorisation of a complex joint distribution into a product of conditional distributions:

\[
p(x_7|x_5, x_6)p(x_6|x_3)p(x_5|x_1, x_2)p(x_4|x_1)p(x_3)p(x_2)p(x_1)
\]
The data

422 Students

10 Topics

10 Tutors

20674 Responses to questions on tests
How was the data collected?

- On Amazon mechanical Turk:
  - ‘Tutors’ are asked to write summaries of Wikipedia articles: these are the ‘lessons’
  - ‘Students’ absorb the summaries, and answer multiple choice questions (10 questions, six answers per question)
Example of the tutor screens

Introduction
This assignment requires summarizing sections from 10 Wikipedia articles.
You will be the tutor of 10 students.
Each of your students will be given the 10 summaries you wrote and 10 questionnaires on the articles. Each questionnaire contains 5 questions.
The bonus for this task depends on how well your students perform. You will be assigned an opponent tutor. The tutor with the best performing students will receive a bonus.
Each summary may contain up to 1500 characters. You will not be shown the questions that your students will be asked; however, the topics of each questionnaire are listed on each topic page in bold.
Summaries that do not include the requested information on the 10 articles will be rejected.
Example questions on the Battle of the Nile

The Battle of the Nile was between:
1. Britain and France
2. Egypt and Israel
3. Egypt and Libya
4. Mamluks and Ottoman Empire
5. Muslims and crusaders
6. Napoleon and the Mamluks

The British forces were led by:
1. Robin Bridge
2. Roy Clare
3. Richard Fitch
4. Edward Harvey
5. Hector Nelson
6. Robert Otway

Click below to continue.

Question 1 of 10

Chad

Read the Wikipedia article on Chad, and summarize basic information about the country, including its location, spoken languages, exports and the history of the civil war.

Please enter your summary:

1500 characters remaining

Screen 1:

Screen 2:
Tutor

- Must summarise each of 10 topics into 1500 words or less
- Does not know what questions will be asked
- Incentive:
  - Gets paid per summary
  - Competes against other tutors for bonuses
Example student screens

Screen 1:

Introduction

You will now be shown a sequence of ten summaries on a variety of topics. Each summary is accompanied by five multiple-choice questions. We would like you to read each summary carefully, and answer the questions as best you can within the time limit specified.

You may find that the summary that you are shown does not contain all the information you need to answer every question on that topic. Don't panic! This is normal. If you don't know the answer, just guess.

Once you have answered the questions, please rate the summary on a scale of 1-5 (with 1 being the lowest and 5 the highest). Once you have answered all the questions and provided a rating, you can move on to the next summary.

When you have completed all the questions, you will be shown the results of your personality test alongside an overview of how well you answered each question.

Click below to continue.

Screen 2:

Chad

Please read the summary below and answer the following questions within the time limit specified. If the summary does not contain all the information required to answer the questions, you can either guess the answer or leave it blank.

Chad is a landlocked country in Central Africa. Libya borders it to the north, Sudan to the east, Central African Republic to the south, Cameroon and Nigeria to the southwest, and Niger to the west. Arabic and French are Chad’s two official languages, and Islam and Christianity are its two most prominent religions. Chad is the seventh poorest country in the world with 50% of its population living in poverty. Cotton is Chad’s dominating export. The Chadian civil war started in 1965 lasting all the way through 1979. The war’s riots and insurgency was against the Chadian president at the time, Francois Tombalbaye.

Where is Chad?

- Central Africa
- East Africa
- North Africa
- North West Africa
- Southern Africa
- West Africa

What is the name of the capital city of Chad?

- Banjul
Student

- For each topic:
  - Is randomly assigned a summary
  - Takes a test of 5 questions – choose one of 6 answers or skip
  - Has limited time to do the test
- Incentive:
  - Gets paid per response
  - Spurious work does not get payment
  - Competes against other students on tests for bonuses
Topics and examples of questions

- Where is Chad?
- Which country participated in Chad’s civil war?
- The length of an adult Willie Wagtail is about?
- A chemical that gives Saffron its taste is?
- Titan's surface temperature is about?
Data analysis and Visualisation is essential for building good models! Can use:
- A general purpose environment such as Matlab or Octave
- A statistical environment such as R
- A spreadsheet
- Home grown tools

When doing data analysis it is useful to have:
- Data in a queryable form
- Good graphing and summarization capabilities
Student/Tutor data set relationships
What problem do we want to solve?

- There may be many...:
  - Evaluate students in a fair way
  - Evaluate tutors in a fair way
  - Assess questions
  - Predict student’s future performance
  - Match students to tutors
Traditional ML versus model-based ML

- In traditional ML, we work towards a specific goal:
  - For example: predict student’s performance
  - Evaluate on a test set using an appropriate metric:
    - For example accuracy, % correct, or log prob(correct)

- In model-based ML, we concentrate on modelling the data:
  - Model evidence is used to assess models
  - Query the model in various ways to solve many problems
  - Use the model to understand and further visualise the data
Result of running Softmax Linear Regression

- Split the data into train and test (14399 train, 6276 test)
- Learn a predictive model using Softmax Linear Regression
  - Inputs:
    - Student, teacher, question, and topic ids
    - Student demographic data
    - Student personality data
  - Labels
    - The student answer (including skip)
- Inputs are bucketised to allow a non-linear mapping to be learned
- Hyper-parameter sweep
- Evaluate using log prob(true label): Training: -0.857, Test: -0.950
## Softmax Linear Regression confusion matrix

<table>
<thead>
<tr>
<th>Truth\Predicted</th>
<th>Skip</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skip</td>
<td>133</td>
<td>43</td>
<td>74</td>
<td>44</td>
<td>24</td>
<td>31</td>
<td>11</td>
<td>0.369</td>
</tr>
<tr>
<td>1</td>
<td>17</td>
<td>1018</td>
<td>120</td>
<td>45</td>
<td>63</td>
<td>41</td>
<td>17</td>
<td>0.771</td>
</tr>
<tr>
<td>2</td>
<td>22</td>
<td>95</td>
<td>1525</td>
<td>52</td>
<td>103</td>
<td>64</td>
<td>4</td>
<td>0.818</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>71</td>
<td>92</td>
<td>520</td>
<td>30</td>
<td>72</td>
<td>22</td>
<td>0.633</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>77</td>
<td>90</td>
<td>53</td>
<td>403</td>
<td>38</td>
<td>34</td>
<td>0.564</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>54</td>
<td>76</td>
<td>34</td>
<td>31</td>
<td>580</td>
<td>8</td>
<td>0.734</td>
</tr>
<tr>
<td>6</td>
<td>13</td>
<td>33</td>
<td>29</td>
<td>44</td>
<td>32</td>
<td>32</td>
<td>220</td>
<td>0.546</td>
</tr>
<tr>
<td>Precision</td>
<td>0.586</td>
<td>0.732</td>
<td>0.760</td>
<td>0.657</td>
<td>0.587</td>
<td>0.676</td>
<td>0.696</td>
<td></td>
</tr>
</tbody>
</table>
Now let’s do some modelling...

Let’s brainstorm what may affect test results...

...suggestions from the floor...
Students

- A student may...
  - systematically guess
  - systematically skip
  - have background knowledge in a topic
  - cheat
  - be careless
  - be impatient
Tutors

- A tutor may
  - be more thorough at preparing summaries than other tutors
  - vary in quality depending on topic
  - get bored after having done a few
  - make mistakes in the summary
Topics and questions

- A topic may
  - may vary in difficulty
  - may vary in question difficulty

- A question may
  - vary in difficulty
  - be unexpected
Some initial visualisations
% correct for student population

Student % correct Histogram

Count of students

% correct

5% 10% 15% 20% 25% 30% 35% 40% 45% 50% 55% 60% 65% 70% 75% 80% 85% 90% 95% 100%

Model-Based Machine Learning Tutorial
Tutor success

Tutor success (99% confidence bars)
Topic success

Topic success (99% confidence bars)
% correct by question

[Bar chart showing percentage correct by question for various categories: Chad, Willie WagTail, Saffron, The Sun Also Rises, DNA, Titan, Armenian Flag, Octopus Card, Zinc, Zhang Heng.]
### Student results per topic and tutor

<table>
<thead>
<tr>
<th>Row Labels</th>
<th>FALSE</th>
<th>TRUE</th>
<th>FALSE</th>
<th>TRUE</th>
<th>FALSE</th>
<th>TRUE</th>
<th>FALSE</th>
<th>TRUE</th>
<th>FALSE</th>
<th>TRUE</th>
<th>FALSE</th>
<th>TRUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>19.61%</td>
<td>80.39%</td>
<td>10.65%</td>
<td>42.49%</td>
<td>57.51%</td>
<td>10.07%</td>
<td>39.89%</td>
<td>60.11%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>38.73%</td>
<td>61.27%</td>
<td>10.66%</td>
<td>41.62%</td>
<td>58.38%</td>
<td>9.04%</td>
<td>24.46%</td>
<td>75.54%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>19.52%</td>
<td>80.48%</td>
<td>10.73%</td>
<td>31.32%</td>
<td>68.68%</td>
<td>9.30%</td>
<td>37.64%</td>
<td>62.36%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>19.42%</td>
<td>80.58%</td>
<td>10.50%</td>
<td>18.27%</td>
<td>81.73%</td>
<td>10.05%</td>
<td>45.40%</td>
<td>54.60%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>46.77%</td>
<td>53.23%</td>
<td>10.28%</td>
<td>11.85%</td>
<td>88.15%</td>
<td>10.79%</td>
<td>35.68%</td>
<td>64.32%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>44.51%</td>
<td>55.49%</td>
<td>9.42%</td>
<td>34.04%</td>
<td>65.96%</td>
<td>9.73%</td>
<td>23.30%</td>
<td>76.70%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>21.13%</td>
<td>78.87%</td>
<td>10.80%</td>
<td>33.33%</td>
<td>66.67%</td>
<td>10.04%</td>
<td>39.78%</td>
<td>60.22%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>44.39%</td>
<td>55.61%</td>
<td>10.33%</td>
<td>27.23%</td>
<td>72.77%</td>
<td>10.06%</td>
<td>37.36%</td>
<td>62.64%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>21.21%</td>
<td>78.79%</td>
<td>10.16%</td>
<td>28.18%</td>
<td>71.82%</td>
<td>9.29%</td>
<td>13.27%</td>
<td>86.73%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>16.90%</td>
<td>83.10%</td>
<td>10.65%</td>
<td>21.57%</td>
<td>78.43%</td>
<td>10.20%</td>
<td>29.80%</td>
<td>70.20%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grand Total</td>
<td>28.86%</td>
<td>71.14%</td>
<td>10.42%</td>
<td>28.62%</td>
<td>71.38%</td>
<td>9.86%</td>
<td>32.30%</td>
<td>67.70%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tutors have radically different outcomes for different topics! Is this because of the student mix in the class?

---

Model-Based Machine Learning Tutorial
A simple model
- it’s those flipping coins again!

α, β

s (Students)
Beta

ability

Bernoulli

q (Question)
correct

correct is a coin flip:

ability is the bias of the coin:

\[
\text{foreach } s \\
a[s] \sim \text{Beta}(\alpha, \beta) \\
\text{foreach } q \\
c[s][q] \sim \text{Bernoulli}(a[s])
\]
Alternative formulation

ability is a Gaussian-distributed ‘score’
correct is T or F depending on whether ability is > or < zero

Ability student 1
Ability student 2

Green area is probability of True
Going beyond a simple ability model

- The DA model
  - Difficulty of a question
  - Ability of the student
Response flowchart for DA model

1. Start
2. Question difficulty
3. Student ability

Triangle:
- Y: Correct
- N: Guess

Flowchart:
- Student ability
- Question difficulty
DA Model for a single response

Pseudo-code:

\[ \text{adv} = a - d \]

if know
\[ \text{sa} = \text{ca} \]
if not know
\[ \text{sa} \sim \text{Discrete}(\pi) \]
Full factor graph for DA model

Definitions:
foreach q
  \( d[q] \sim \text{Gaussian}(...) \)

foreach s
  \( a[s] \sim \text{Gaussian}(...) \)

Response loop:
foreach r
  \( \text{adv} = a[s[r]] - d[q[r]] \)
  if know
    \( sa[r] = ca[q[r]] \)
  if not know
    \( sa[r] \sim \text{Discrete}(\pi) \)
DA model results

![Graphs showing model results.

- % correct versus ability
  - Avg. Difficulty: -0.84
  - Avg. Difficulty: -0.48

- % correct versus difficulty
  - Where is Chad?
  - Easy question not covered
  - In one lesson.
## DA model predictions (training set)

<table>
<thead>
<tr>
<th>Truth\Predicted</th>
<th>Skip</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skip</td>
<td>0</td>
<td>188</td>
<td>247</td>
<td>107</td>
<td>132</td>
<td>156</td>
<td>54</td>
<td>0.000</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>2324</td>
<td>216</td>
<td>191</td>
<td>160</td>
<td>124</td>
<td>21</td>
<td>0.765</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>221</td>
<td>3603</td>
<td>103</td>
<td>288</td>
<td>122</td>
<td>24</td>
<td>0.826</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>117</td>
<td>203</td>
<td>1266</td>
<td>150</td>
<td>170</td>
<td>42</td>
<td>0.650</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>167</td>
<td>166</td>
<td>98</td>
<td>1057</td>
<td>87</td>
<td>33</td>
<td>0.657</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>100</td>
<td>127</td>
<td>101</td>
<td>97</td>
<td>1293</td>
<td>19</td>
<td>0.744</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>67</td>
<td>63</td>
<td>108</td>
<td>149</td>
<td>74</td>
<td>364</td>
<td>0.441</td>
</tr>
<tr>
<td><strong>Precision</strong></td>
<td>0.000</td>
<td>0.730</td>
<td>0.779</td>
<td>0.641</td>
<td>0.520</td>
<td>0.638</td>
<td>0.654</td>
<td></td>
</tr>
</tbody>
</table>

Model’s not able to capture ‘skip’ as a response
(... though this table is just based on mode of prediction)
Let’s go back to the data

214 students never skip
68 student skip once
DASG Model

- We consider two extensions to the DA model
  - We explicitly model the skip action of a student
  - We learn the bias in the guess probabilities
Response flowchart for DASG model

Start

- Skipped? (Y → Skip, N → Know?)
  - Propensity to skip
  - Student ability
  - Question difficulty

Know? (Y → Correct, N → Guess)

Correct

Guess
DASG Model for a single response

Model-Based Machine Learning Tutorial
How do we know this model is better?

- Compare model log evidence:

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>DA</td>
<td>-14480</td>
</tr>
<tr>
<td>DAS</td>
<td>-14263</td>
</tr>
<tr>
<td>DASG</td>
<td>-14062</td>
</tr>
</tbody>
</table>

- A value of 5 represents significant increase in evidence
- Both the skip variable, and learning guess probabilities separately give a significant improvement
Comparing models with model evidence

- Probability of the data with all parameter uncertainty integrated out:
  \[ p(D|M) = \int_\theta p(D|\theta) p(\theta|M) d\theta \]

- Compare two models using evidence:
  \[ p(D|M_1) \geq p(D|M_2) ? \]

- Normally work in terms of log evidence:
  \[ \log(p(D|M_1)) - \log(p(D|M_2)) > 4.6 \Leftrightarrow \frac{p(D|M_1)}{p(D|M_1) + p(D|M_2)} > 0.99 \]
DASG model results

Answered every question
Skipped 22 answers
Cheats?

% correct versus ability

% correct versus propensity to skip

Inferred Ability

Inferred propensity to skip

Non-skippers
Skippers

Model-Based Machine Learning Tutorial
# DASG model predictions (training set)

<table>
<thead>
<tr>
<th>Truth \ Predicted</th>
<th>Skip</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skip</td>
<td>137</td>
<td>168</td>
<td>221</td>
<td>83</td>
<td>81</td>
<td>143</td>
<td>51</td>
<td>0.155</td>
</tr>
<tr>
<td>1</td>
<td>14</td>
<td>2319</td>
<td>255</td>
<td>168</td>
<td>140</td>
<td>119</td>
<td>21</td>
<td>0.764</td>
</tr>
<tr>
<td>2</td>
<td>32</td>
<td>218</td>
<td>3630</td>
<td>96</td>
<td>254</td>
<td>110</td>
<td>21</td>
<td>0.832</td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>116</td>
<td>238</td>
<td>1254</td>
<td>130</td>
<td>158</td>
<td>40</td>
<td>0.644</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
<td>161</td>
<td>184</td>
<td>92</td>
<td>1038</td>
<td>86</td>
<td>32</td>
<td>0.646</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
<td>99</td>
<td>143</td>
<td>97</td>
<td>84</td>
<td>1286</td>
<td>19</td>
<td>0.740</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>67</td>
<td>73</td>
<td>97</td>
<td>144</td>
<td>74</td>
<td>360</td>
<td>0.436</td>
</tr>
<tr>
<td>Precision</td>
<td>0.598</td>
<td>0.737</td>
<td>0.765</td>
<td>0.665</td>
<td>0.555</td>
<td>0.651</td>
<td>0.662</td>
<td></td>
</tr>
</tbody>
</table>
DASG guess probabilities

- Top-heavy bias in the guessing!
What about cheating?

- Some questions are covered rarely or not at all, and are difficult to guess
  - Titan's surface temperature is about?
  - The boiling point of zinc is?
  - Which of the following molecules are building blocks for DNA?
What about learning?

- So far we have not looked at the lessons
- These should have a big effect – the lesson is in front of the student!
  - How thorough has the tutor been in preparing the lesson?
  - How diligent/persevering has the student been in reading the lesson?
DASGCL model

- Adds two more concepts
  - Cheating – a student may decide to cheat on the tests
  - Lesson/Learning
    - Each tutor has a thoroughness variable
      - This is used to condition a ‘covered’ variable (per question per tutor)
    - Each student has a diligence variable
DASGCL Model for a single response

- Student
- Tutor
- Question
- Response

Model-Based Machine Learning Tutorial
Is the model an improvement?

<table>
<thead>
<tr>
<th>Model</th>
<th>Log Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>DA</td>
<td>-14480</td>
</tr>
<tr>
<td>DAS</td>
<td>-14263</td>
</tr>
<tr>
<td>DASG</td>
<td>-14062</td>
</tr>
<tr>
<td>DASGC</td>
<td>-14021</td>
</tr>
<tr>
<td>DASGCL</td>
<td>-12883</td>
</tr>
</tbody>
</table>

Yes! Much greater than our rule of thumb value of 5.
DASGCL – cheating

% correct versus ability

% correct

Inferred ability

Extroversion versus cheat

0 1 2 3 4 5 6

Cheaters Non-Cheaters

40 of them

Model-Based Machine Learning Tutorial
DASGCL – thoroughness and coverage

Inferred versus actual coverage:
## Inferred versus Actual Coverage

<table>
<thead>
<tr>
<th>True\Inferred</th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>175</td>
<td>9</td>
</tr>
<tr>
<td>Part</td>
<td>19</td>
<td>15</td>
</tr>
<tr>
<td>Yes</td>
<td>43</td>
<td>239</td>
</tr>
</tbody>
</table>
DASGCL: Difficulty

% correct versus difficulty

67% correct
Difficulty: 1.5
Coverage: 70%

63% correct
Difficulty: 0.05
Coverage: 28%

Inferred question difficulty
# DASGCL model predictions (training set)

<table>
<thead>
<tr>
<th>Truth</th>
<th>Predicted</th>
<th>Skip</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Recall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skip</td>
<td>Skip</td>
<td>317</td>
<td>106</td>
<td>217</td>
<td>62</td>
<td>43</td>
<td>98</td>
<td>41</td>
<td>0.36</td>
</tr>
<tr>
<td>1</td>
<td>36</td>
<td>2294</td>
<td>450</td>
<td>78</td>
<td>83</td>
<td>79</td>
<td>16</td>
<td></td>
<td>0.76</td>
</tr>
<tr>
<td>2</td>
<td>53</td>
<td>152</td>
<td>3873</td>
<td>69</td>
<td>125</td>
<td>73</td>
<td>16</td>
<td></td>
<td>0.89</td>
</tr>
<tr>
<td>3</td>
<td>28</td>
<td>95</td>
<td>405</td>
<td>1232</td>
<td>53</td>
<td>105</td>
<td>30</td>
<td></td>
<td>0.63</td>
</tr>
<tr>
<td>4</td>
<td>23</td>
<td>137</td>
<td>289</td>
<td>63</td>
<td>1006</td>
<td>65</td>
<td>25</td>
<td></td>
<td>0.63</td>
</tr>
<tr>
<td>5</td>
<td>17</td>
<td>82</td>
<td>238</td>
<td>76</td>
<td>47</td>
<td>1263</td>
<td>14</td>
<td></td>
<td>0.73</td>
</tr>
<tr>
<td>6</td>
<td>23</td>
<td>55</td>
<td>210</td>
<td>68</td>
<td>61</td>
<td>50</td>
<td>358</td>
<td></td>
<td>0.43</td>
</tr>
<tr>
<td>Precision</td>
<td>0.64</td>
<td>0.79</td>
<td>0.68</td>
<td>0.75</td>
<td>0.71</td>
<td>0.73</td>
<td>0.72</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
What about personality and demographics data?

- We can attach linear regression prior to any of our ‘score’ latent variables:
- Bucketise real-valued features to allow nonlinearity
Age analysis
Encoding age

- Bucketisation allows learning non-linear relationships
- Create buckets at regularly spaced quantiles (‘nodes’)
- Values in buckets represent how close to each node
  - Use linear interpolation or kernel

- Nodes at ages: 15, 23.5, 26, 31, 38, 68.
- Using linear interpolation, the value 30 is encoded as:
  \((0.0, 0.0, 0.2, 0.8, 0.0)\)
Personality analysis

- Openness
- Conscientiousness
- Extroversion
- Agreeableness
- Neuroticism/Emotional Stability
Encoding Personality

- **Extroversion** is well represented across scores – encode as one bin per score
- Other personality traits are poorly represented at low scores
  - Consider merging low scores into one bucket
  - Use cumulative histogram to decide how to merge scores into the same bucket:

```
Bin 1: Score 0 or less
Bin 2: Score 1 or 2
Bin 3: Score 3
Bin 4: Score 4
Bin 5: Score 5
Bin 6: Score 6
```
DASGCLF model

- Adds features by attaching linear regression priors on score variables
- We just show ‘Extroversion’ and ‘languageIsEnglish’ attached to the ‘diligence’ latent variable.
DASGCLF Learned weights

### Model Log Evidence

<table>
<thead>
<tr>
<th>Model</th>
<th>Log Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>DASGCL</td>
<td>-12883</td>
</tr>
<tr>
<td>DASGCLF</td>
<td>-12857</td>
</tr>
</tbody>
</table>
Evidence versus log prob.(true label)

DA: Difficulty/ability  
DAS: DA + skip  
DASG: DAS + guess probabilities  
DASGC: DASG + cheating  
DASGCL: DASGC + learning  
DASGCLF: DASGCL + features  
MLR: Multiple Linear Regression
Evidence versus log-prob(true label)

- Curves for evidence and log-prob(true label) are very similar.

- Once enough data has been seen, evidence measures the performance on an online predictor

\[
\log p(r_1, r_2, r_3, \ldots, r_n) = \\
\log p(r_1) + \log p(r_2 | r_1) + \log p(r_3 | r_2, r_1) \\
+ \ldots + \log p(r_n | r_{n-1}, \ldots, r_1)
\]
Thanks...

- ... to Tom Minka and the Infer.NET team for help with the modelling
- ... to Mike Armstrong and Yoram Bachrach for help with the data

If you are interested in this tutorial data, email:
To: joguiver – at – microsoft.com
Subject: MBML tutorial