Problem Statement

This study applies machine learning to this new data source to understand how investing in tutoring products change students’ learning trajectories. Machine learning techniques are used to model student’s progress with student-question interactions as inputs and predicted achievement (operationalized as probability of a correct answer across all questions) as output.

Motivation

Online education promise to revolutionize education but companies struggle to show that their products improve student learning. To tackle this problem, I seek to leverage machine learning to estimate student learning.

Dataset

Data for this project comes from a mobile education company in China. The data contains logs of student questions responses on homework assigned by their school teacher. For the estimation, I use 138,001 records of student question responses from one school in one month in 2018.

Models

- Naive Bayes
  \[
  \Pr(y_i = 1; t) = \frac{\sum_j \gamma(t_{ij} - t) \cdot \mathbb{1}\{y_{ij} = 1\}}{\sum_j \gamma(t_{ij} - t) \cdot \mathbb{1}\{i \text{ attempted } j\}}
  \]
  where \(\gamma(\cdot)\) is a Gaussian kernel.

- Difficulty-Modulated Naive Bayes (with Laplace smoothing)
  \[
  \hat{\theta}_j = \frac{\sum_i \mathbb{1}\{y_{ij} = 1\}}{\sum_i \mathbb{1}\{i \text{ attempted } j\}}
  \]
  \[
  \hat{\Pr}(y_i = 1) = \frac{\sum_j \gamma(t_{ij} - t) \mathbb{1}\{y_{ij} = 1\}/\hat{\theta}_j}{\sum_j \gamma(t_{ij} - t) \mathbb{1}\{i \text{ attempted } j\}}
  \]

- Siamese-Like Neural Network

Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Class Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F1 Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Siamese</td>
<td>0.873</td>
<td>0.907</td>
<td>0.937</td>
<td>0.922</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Batch Size</th>
<th>Learning Rate</th>
<th>Epochs</th>
<th>Network Structure</th>
<th>Feature Vector Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>128</td>
<td>0.001</td>
<td>15</td>
<td>Student: [400-80] Question: [200-80]</td>
<td>10</td>
</tr>
</tbody>
</table>