Feature Engineering for Clustering Student Solutions
Elena Glassman, Rishabh Singh, Rob Miller
MIT CSAIL
(elg.rishabhrcm)@mit.edu

Problem
- There may be several distinct, correct solutions to coding assignments.
- Some solutions may be unknown to the teaching staff or intelligent tutor designer.
- This complicates the task of providing help and hints.

Approach
- **Visualize** hundreds or thousands of student solutions to find alternatives.
- **Classify** solutions by their design choices
  - By machine learning and/or human staff
  - Use this knowledge to enhance
  - Activities about design choices and tradeoffs
  - Peer-pairing
  - Automated help

Case Study 1: Turing Machine (TM) Synthesis Lab @ MIT
- Most staff knew of one TM solution.
- Dynamic behavior of TM solutions shows two solution clusters.

Case Study 2: Matlab Programming Game Visualization
- This interactive visualization allows users to explore and compare other submitted solutions.

Work In Progress: Visualizing Code Variation in Python for Insight and Feature Discovery
- Joint work with Ned Gulley (MathWorks)

---

### Table: Implementations that pair

<table>
<thead>
<tr>
<th>Problem Statement</th>
<th>Initial TM Code</th>
<th>Student-written TM code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Function asked for by teacher</td>
<td>Initial TM Code</td>
<td>Student-written TM code</td>
</tr>
<tr>
<td>Different function</td>
<td>Initial TM Code</td>
<td>Student-written TM code</td>
</tr>
<tr>
<td>What features are useful for visualization?</td>
<td>Initial TM Code</td>
<td>Student-written TM code</td>
</tr>
</tbody>
</table>

### Diagrams:

- Feature Engineering for Clustering Student Solutions
- Turing Machine Input Tape
- High-Level and Low-Level differences
- Separation dimension
- Interaction visualization
- Joint work with Ned Gulley (MathWorks)