Applying Machine Learning to the Assessment of Problem-Solving Skills

Max Arseneault marsenea@stanford.edu

EDUCATION

Introduction

This project explores how machine learning algorithms could be applied to assess college students' scientific problem-solving skills using log data generated in an interactive circuit simulation.

Specifically, **we investigated the performance of different machine learning algorithms using sequences of students' interactions as features** to predict their problem-solving performance as measured by the solution scores.

Our first attempts at modeling were void of artificial intelligence techniques. Mediocre results motivated attempts to learn solution scores with an artificial neural network as well as with dimensionality reduction and clustering techniques.

Data

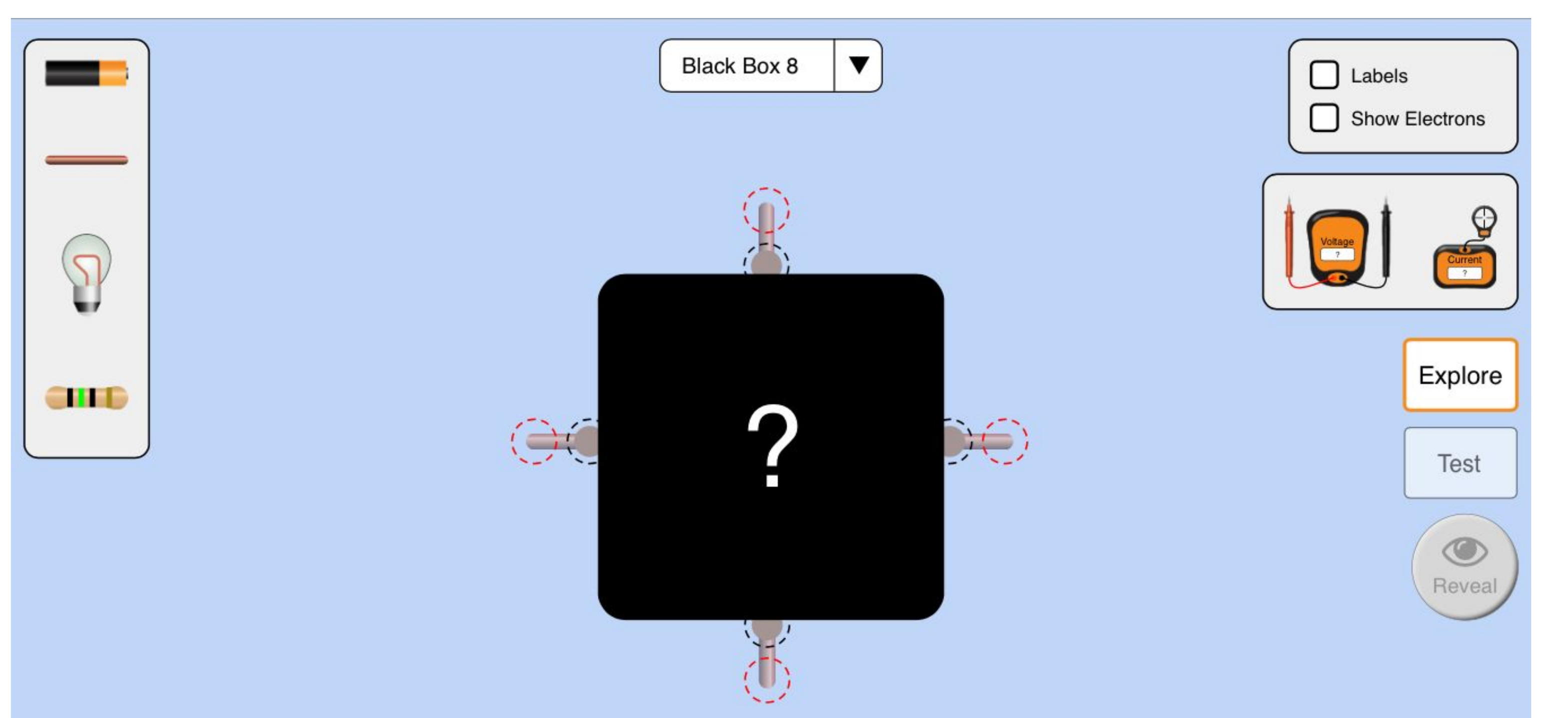
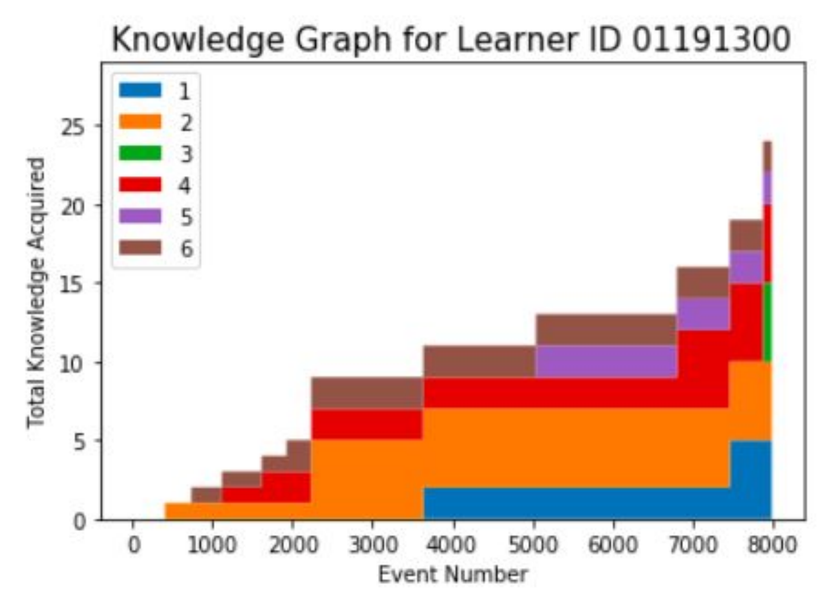
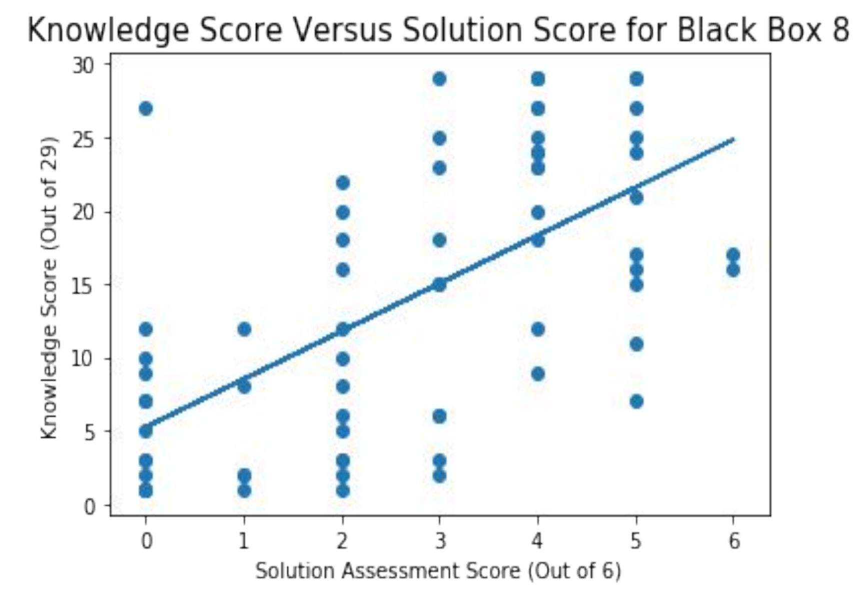


Figure 1. Illustration of the Black Box Simulation

- A group of college students participated in the study using the PhET Circuit Construction Kit Black Box (*Figure 1*), yielding a valid dataset of 178 samples
- Log files of individual participants' interactions were parsed into a sequence of time-stamped *events*
- The predicted variable is participants' problem-solving performance as measured by the solution score (0 - low performing, 1 - high performing)

Non-AI Attempts





Features Per Loop (x6)

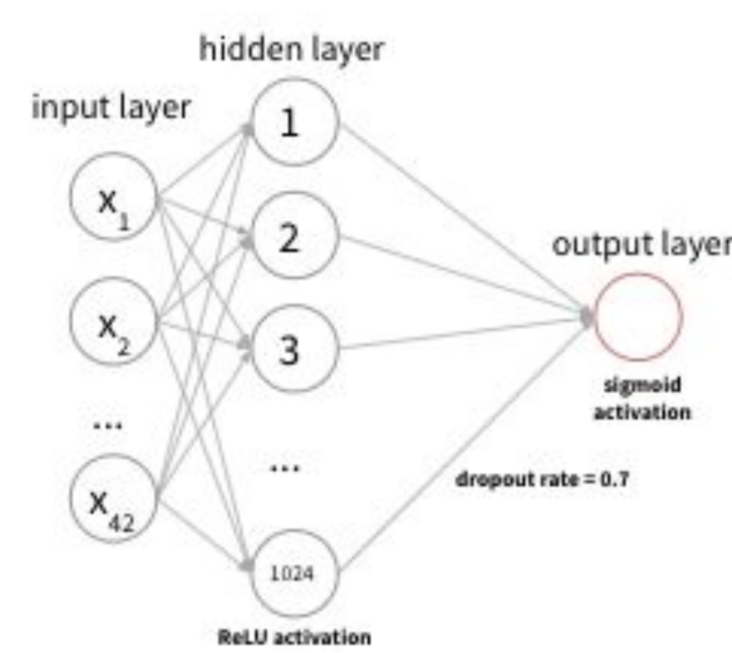
Feature	Description
battery & lightbulb	constructed loop with battery & lightbulb
battery & resistor	constructed loop with battery & resistor
battery	constructed loop with only a battery
lightbulb	constructed loop with only a lightbulb
resistor	constructed loop with only a resistor
wire	constructed loop only with wire
voltmeter	performed a valid voltmeter measurement

Top-Performing Model

Neural Network

Input layer - 42 features
Hidden layer - 1024 units
Output layer - 1 unit
  0: low problem-solving skill
  1: high problem-solving skill
Cost function: Binary cross-Entropy

$$BCE = -\frac{1}{N}\sum_{i=0}^{N} y_i \cdot log(\hat{y}_i) + (1-y_i) \cdot log(1-\hat{y}_i)$$



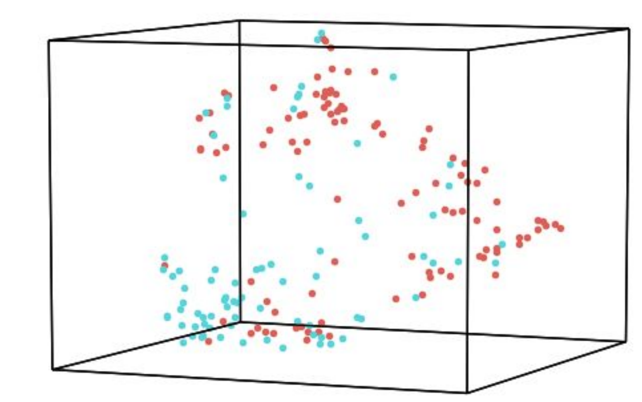
Results

Model	Training Accuracy	Test Accuracy	Precision (Class 0/1)	Recall (Class 0/1)	F1 Score
Neural Network	88.8%	75.7%	77.8%	87.5%	0.824

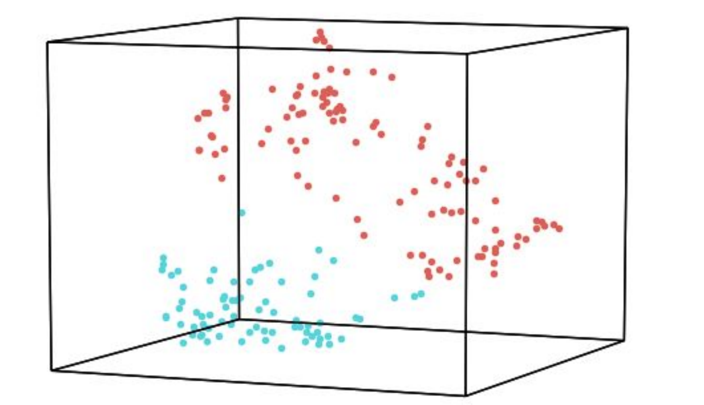
$$Precision = \frac{TP}{TP+FP} \quad Recall = \frac{TP}{TP+FN} \quad F1 = \frac{2*Precision*Recall}{Precision + Recall}$$

Unsupervised Learning

UMAP Dimension Reduction



True Labels



K-Means (k = 2)

Discussion & Future

- Our neural network approach achieved **high training and test accuracies with a small train-test gap**. Performance on other traditional metrics also far exceeded human-level performance.
- Our efforts at unsupervised learning are promising and hint at latent structure in the dataset

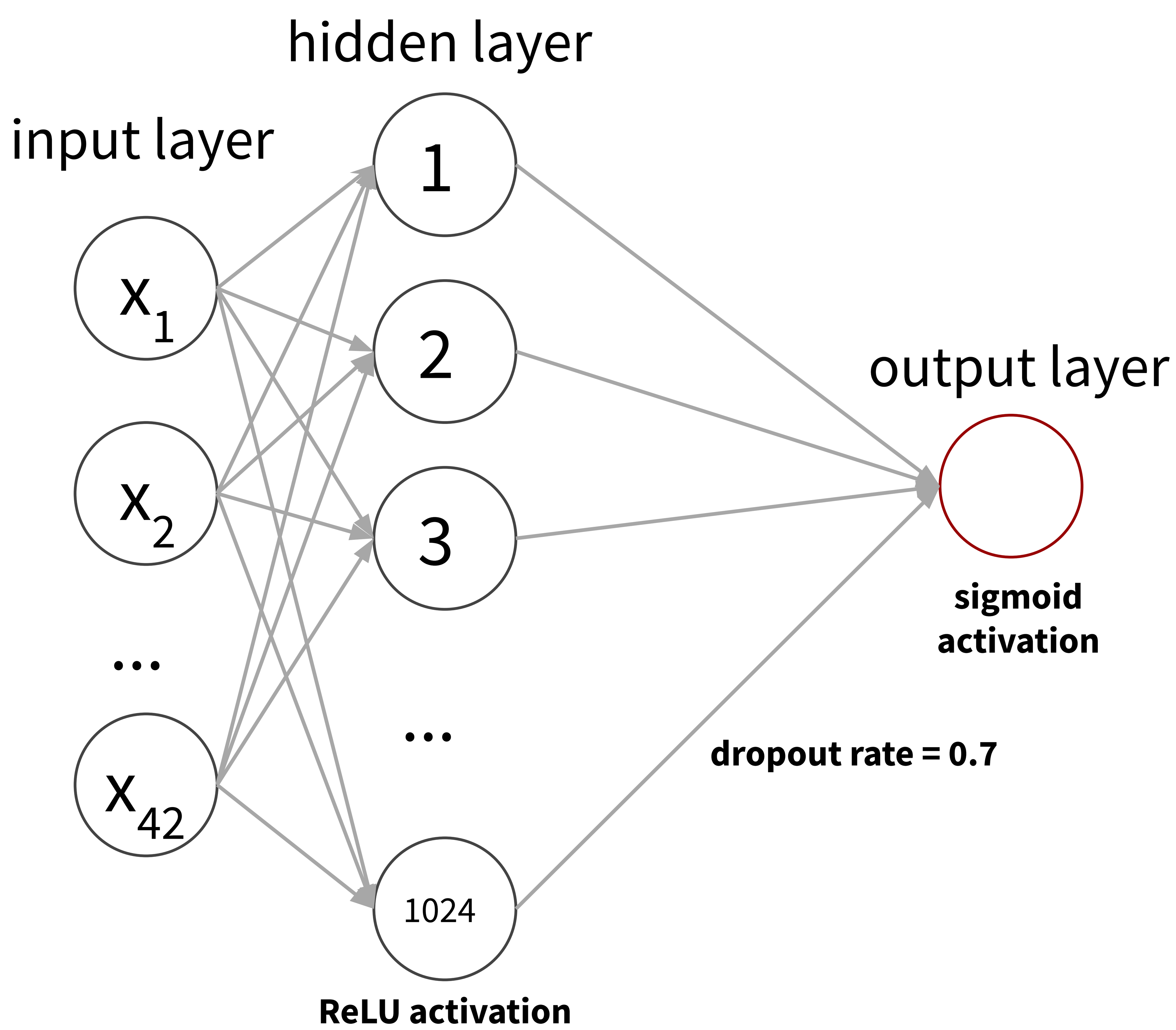
Reference & Acknowledgements

[1] Piech, Chris, et al. "Deep knowledge tracing." *Advances in neural information processing systems*. 2015.

The project received advice and guidance from Professor Carl Wieman at the Graduate School of Education and Department of Physics, and Karen Wang, a research student in the Wieman Group.

PhET INTERACTIVE SIMULATIONS

GORDON AND BETTY MOORE FOUNDATION


hidden layer
input layer
output layer
x1
x2
...
x42
1
2
3
...
1024
ReLU activation
sigmoid
activation
dropout rate = 0.7