DISSERTATION

by

Geoffrey Converse

A thesis submitted in partial fulfillment
of the requirements for the
Doctor of Philosophy
degree in Applied Mathematical And Computational Sciences in the
Graduate College of
The University of Iowa

December 2021

Thesis Supervisor: Professor Suely Oliveira

ACKNOWLEDGEMENTS

Prior to your first thesis deposit, replace this text with your acknowledgements. This text should be double spaced and each paragraph should be indented. This text may be altered between first and final deposit.

Prior to your first thesis deposit, replace this text with your acknowledgements. This text should be double spaced and each paragraph should be indented. This text may be altered between first and final deposit.

ABSTRACT

***

Prior to your first thesis deposit, replace this text with the text of your scientific/ scholarly abstract. The text of this abstract should be double spaced and each new paragraph should be indented. This text may be altered between first and final deposits.

This abstract is required for everyone except DMA and MFA students.

PUBLIC ABSTRACT

***

Prior to your first thesis deposit, replace this text with the text of your public abstract. The text of this abstract should be double spaced and each new paragraph should be indented. The text may be altered between first and final deposits. This abstract is required for all thesis/dissertations.

The public abstract is to be placed at this point in your first and final deposit and submitted via webform at final deposit. This abstract may be up to 250 words and should be written for a non-academic lay audience. In writing your public abstract, avoid jargon and technical language as much as possible.

The ability to communicate research simply and clearly is an important skill when interviewing for faculty positions, as well as for positions in industry and alt-ac sectors. The public abstract helps convey ideas beyond one's immediate academic circle, facilitating communication with colleagues who do different kinds of work and possess different dimensions of training.

Think of your public abstract as your "elevator pitch" or what you might tell someone who asks, "What is your thesis about?" You may only have a few minutes to explain it to them while keeping their attention and using terminology you are sure they will understand without further lengthy explanation.

Another way to think of your public abstract is like the description you would read on the inside of a book cover.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

Prior to your first deposit, replace this text with the text of your Preface. The Preface should be double spaced and new paragraphs should be indented. This text may be altered between first and final deposits.

THIS PAGE IS OPTIONAL

BACKGROUND - IRT PARAMETER ESTIMATION

In educational measurement, a common goal is to quantify the knowledge of students from the results of some assessment. In a classroom setting, grades are typically assigned based on the percentage of questions answered correctly by a student assignments. The letter grades assigned from these percentages can serve as a naive measure of student knowledge; "A" students have completely mastered the material, "B" students have a good grasp of material, "C" students are fairly average, and "D" and "F" students have significant gaps in their knowledge.

The practice of evaluating student ability purely from a raw percentage score is known as true score theory [thissen] [thissen] thissen. But there are clear issues with this approach. Not all questions on an exam or homework assignment is created equally: some questions are easier, and some more difficult. Consider a scenario where two students both answer 17 out of 20 questions correctly on a test for a raw score of 85%. But if Student A answered questions 1, 8, and 9 wrong while Student B answered 4, 17, and 20 incorrectly, it is not likely that that Student A and Student B possess the same level of knowledge. For example, questions 1, 8, and 9 could be much more difficult than questions 4, 17, and 20. Additionally, the two sets of problems could cover different types of material. True score theory does not account for either of these situations, and naively quantifies the knowledge of Student A and Student B as equal.

More sophisticated methods have been studied which attempt to more accurately quantify student learning. Cognitive Diagnostic Models (CDM) (TODO: citation) aim to classify whether students possess mastery of a given skill or not. This discrete classification can be useful in determining whether or not a student meets a prerequisite, or deciding whether or not they are ready to move on to the next level of coursework. We focus instead on Item Response Theory, where student knowledge is assumed to be continuous.

Item Response Theory

Item Response Theory (IRT) is a field of quantitative psychology which uses statistical models to model student ability. These models often give the probability of a question being answered correctly as a function of the student's ability. In IRT, it is assumed that each student, indexed by j, possesses some continuous latent ability θ_j. The term "latent ability" is synonymous with "knowledge"or "skill." Often, it is assumed that amongst the population of students, $\theta_j \sim \mathcal{N}(0,1)$ [thissen].

In this work, we often consider the case where each student has multiple latent abilities. For example, in the context of an elementary math exam, we may wish to measure the four distinct skills "add", "subtract", "multiply", and "divide." This scenario is referred to as multidimensional item response theory, and we write the set of student j's K latent abilities as a vector $\Theta_j = (\theta_{1j}, \theta_{2j}, \ldots, \theta_{Kj})^\top$. It is then assumed that the latent abilities of students follow some multivariate Gaussian distirbution, $\mathcal{N}(0, \Sigma)$. For simplicity, the covariance matrix Σ is often taken to be the identity matrix, making each latent skill independent of one another.

Note that Θ_j is not directly observable in any way. Instead, a common goal is to infer student's knowledge Θ_j from on their responses on some assessment containing n questions, referred to as items. A student's set of responses can be written as a binary n-dimensional vector $U_j = (u_{1j}, u_{2j}, \ldots, u_{nj})^\top$, where

$$u_{ij} = \begin{cases} 1 & \text{if student } j \text{ answers item } i \text{ correctly} \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

IRT models aim to model the probability of a student answering a particular question correctly, so that the probability of student j answering item i correctly is given by some function of Θ_j:

$$P(u_{ij} = 1 | \Theta_j) = f(\Theta_j; V_i) \tag{2}$$

where V_i is a set of parameters associated with item i. In general, $f : \mathbb{R}^K \to [0,1]$ is some continuous function which is strictly increasing with respect to Θ_j.

In the next three sections, we describe various candidates for the function f. Though each is presented in the context of single-dimensional IRT ($K = 1$), they can all be easily adapted to higher dimensions.

Rasch Model

Normal Ogive Model

2-parameter Logistic Model

Parameter Estimation Methods

Artificial Neural Networks

In recent years, artifical neural networks (ANN) have become an increasingly popular tool for machine learning problems. Though they have been around since the 1960's (TODO: citation), GPU technology has become more accessible and modern computers are more powerful, allowing anyone interested to train a basic neural network on their machine. ANN can be applied to a diverse set of problems, including regression, classification, computer vision, natural language processing, function approximation, data generation, and more (TODO: citations).

One of the biggest critiques of ANN is their black-box nature, meaning that the decision process that a trained model uses is typically not explainable by humans. As opposed to simpler methods such as decision trees or linear regression, neural networks are not interpretable. This makes them less desirable in certain applications where researchers wish to know why a model predicts a particular data sample the way that it does. For example, if a financial institution is using data science methods to determine whether or not to approve someone's loan, the institution should be able to explain to the customer why they were denied. Most customers will not be satisfied with "the computer told us so," and there is a possibility that a black-box neural network could learn and use features such as race or gender in its prediction, which is illegal in the United States (TODO: definitely need citation or delete).

Autoencoders

Variational Autoencoders

METHODS - IRT PARAMETER ESTIMATION

Ml2p-vae Description

One-parameter Logistic

2-parameter Logistic

3-parameter Logistic

tbd on this

Full Covariance Matrix Implementation

Software Package In R Description

Plug that I made this and it is publicly available
Example of a table:

Table 1. My first table.

column	column	column	column	column	column	column	column	column	column
12.34	12.34	12.34	12.34	12.34	12.34	12.34	12.34	12.34	12.34
12.34	12.34	12.34	12.34	12.34	12.34	12.34	12.34	12.34	12.34

RESULTS - IRT PARAMETER ESTIMATION

Description of Data Sets

1-pl Results

2-pl Results

3-pl Results

(maybe)

KNOWLEDGE TRACING BACKGROUND

Application Goal

Mathematical Setup

Literature Review

Bayesian Knowledge Tracing

Deep Knowledge Tracing

Dynamic Key-value Memory Networks

KNOWLEDGE TRACING - METHODS

Item-based Attention Networks

TODO:name this better

KNOWLEDGE TRACING - RESULTS

Data Description

Describe each dataset used here.

Experiment Details

Hyper parameters here

Results