Research Experiences For Undergraduates

UCLA Department of Mathematics

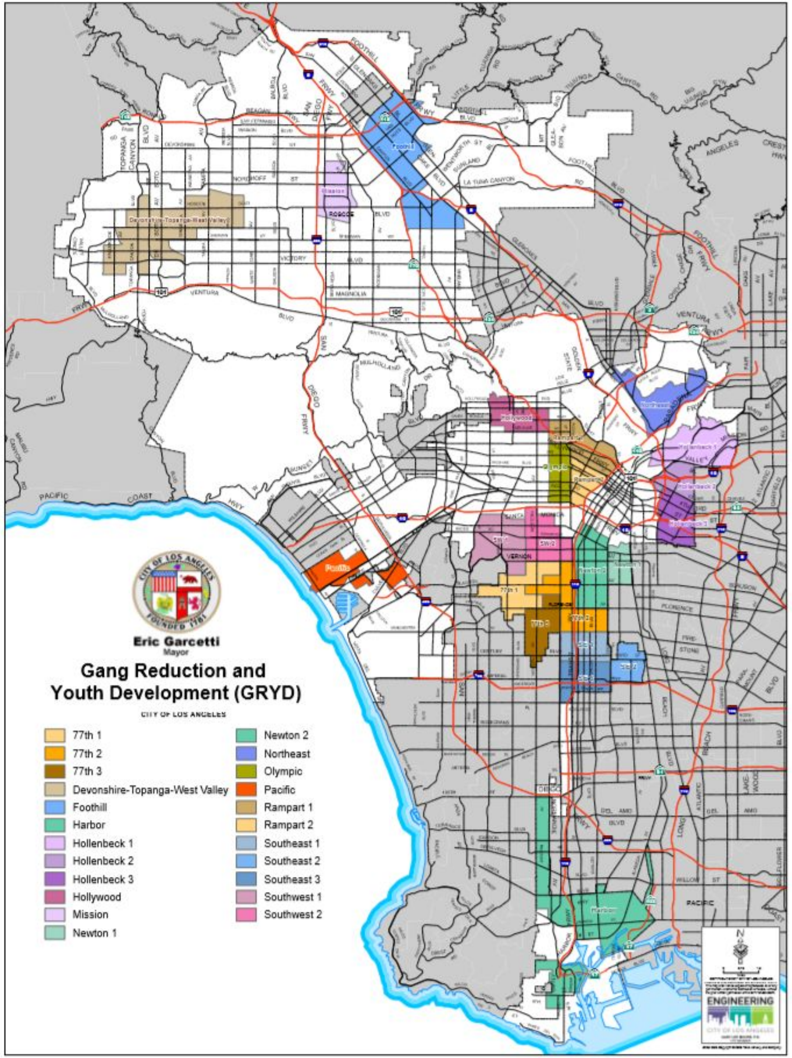
Gang Reduction & Youth Development Project (GRYD)

Zehan Chao, Zheyuan (Kevin) Cui, Avery Edson, Cesar Guajardo, Yihuan (Charlotte) Huang, Xingjia Wang, Zhanyuan (Jerry) Yin

Mentors: Dr. Heather Z. Brooks, Dr. Hanqin Cai

What is GRYD?

- Program conducted by the City of Los Angeles Mayor's Office
- Aims to:
  - **Reduce** and curb **gang violence**
  - Promote **development** for at-risk youth
- 23 GRYD Zones throughout LA



(**Y**outh **S**ervices **E**ligibility **T**ool)

YSET

Questionnaire

56 / **104** Questions

9 / **12** Sections

Attitudinal
(39 observed)

SCALE F	Do you agree or disagree with these statements? SHOWCARD 3	strongly agree	agree	neither agree nor disagree	disagree	strongly disagree
F21	It is okay for me to lie (or not tell the truth) if it will keep my friends from getting in trouble with parents, teachers or police.	5	4	3	2	1

Behavioral
(17 observed)

SCALE IJ — **People sometimes break rules or laws. I'd like you to be honest with me about the rules or laws you have broken in your entire life and in the last six months. Remember, your answers will stay private.**

	Have you . . . SHOWCARD 6	a. In the last 6 months		b. EVER		c. With gang	
IJ40	Used alcohol or cigarettes?	N	Y	N	Y	N	Y

YSET

Dataset

1. **Data Cleaning**
   - > Construct Identifiers
   - > Check consistency
   - > Assign NA to invalid values
2. **Exploratory Analysis**
   - > Correlation between variables
   - >Factor Analysis/ Dimension Reduction
3. **Variable Selection**
   - > Reconstruction of variables
   - > Select variables according to areas of interest

32896 observations

22567 participants

~1600 participants in YRR Dataset

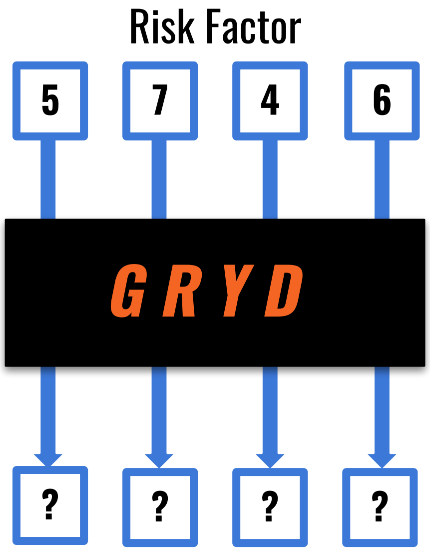
Administrative
- > *UniqueID*
- > *Intake/Retake*
- > *Date*

Demographic
- > *Age*
- > *Gender*
- > *Ethnicity*

Response
- > *By question*
- > *By section*
- > *By concern lv.*

Risk Factor
- > *Eligibility*
- > *Original RF*
- > *Reconstructed RF*

Areas of Interest



1. Evaluating the change in responses
2. Importance of questions in calculating the risk factor
3. Prediction of future “risk variables”
   (Risk Factor, Eligibility, Concern Levels, etc)

How effective is the GRYD program?

Agenda

Introduce methods to evaluate:
1. Change in responses
2. Importance of questions

Introduction

Understanding
Change vs Importance

Interpretation

Examination of prediction accuracy on test sets

Examination

Verify methods by comparing with **machine learning** algorithms

Comparison

Future work

Discussion

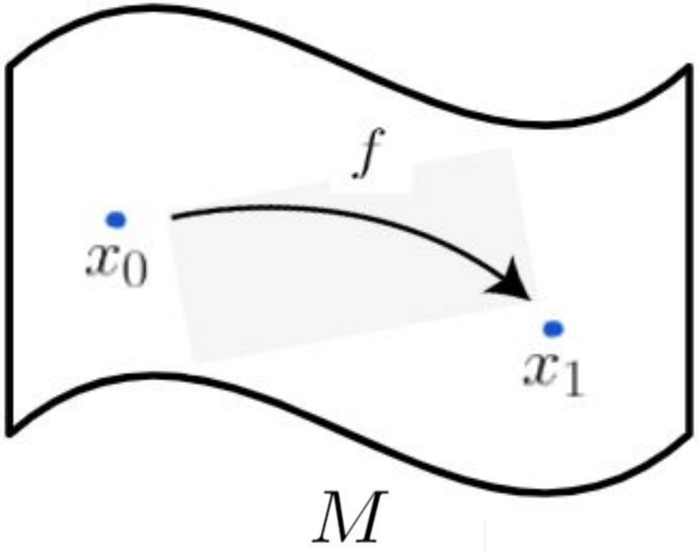
Dynamical System

"Participant points" on manifold:

$$x_k \in M$$

Dynamics of the system:

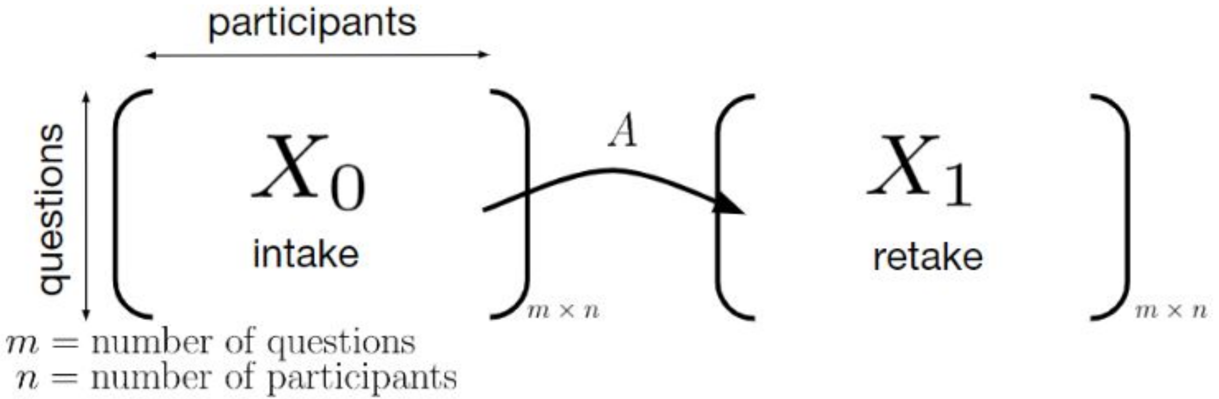
$$x_{k+1} = f(x_k)$$



Dynamic Mode Decomposition

DMD

Algorithm



1) Find linear transformation matrix **A** such that **AX_0** approximates **X_1**: $\| AX_0 - X_1 \|_F \approxeq 0$

2) Compute the Singular Value Decomposition of **X_0**: $X_0 = U\Sigma V^*$

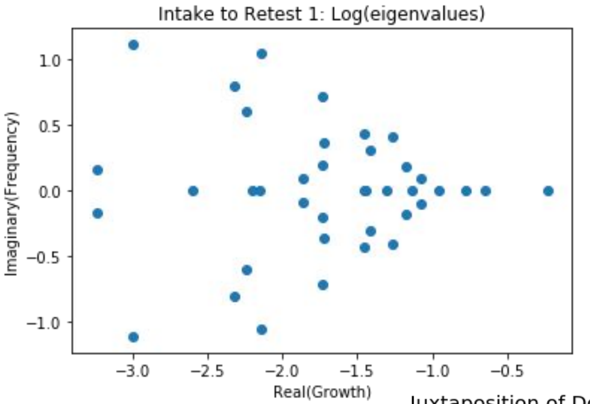
3) Compute the transformation matrix **A**: $A = (X_1)(X_0)^\dagger = X_1 V \Sigma^{-1} U^*$

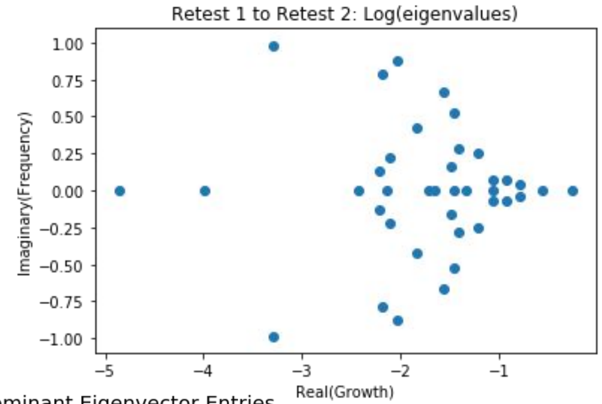
4) Compute dominant eigenvalues and eigenvectors of **A**

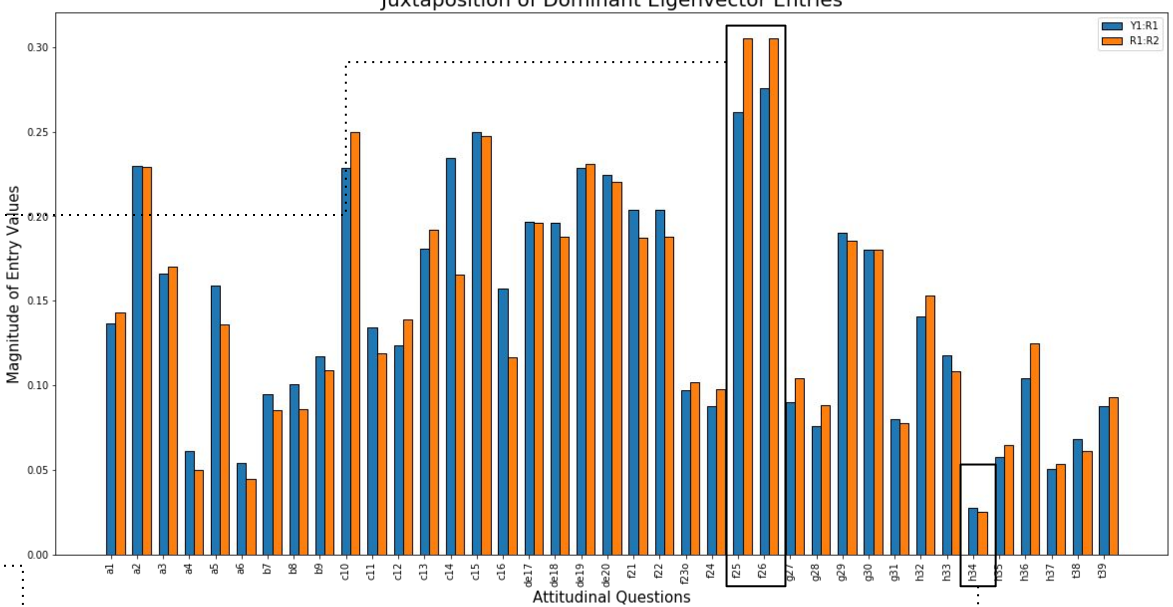
DMD

Results

(Questions' susceptibility to **change**)







***F25**: It is okay to beat people if they beat me first.*
***F26**: It is okay to beat people if I do it to stand up for myself.*

are reluctant to change

***H34**: How many of your friends have attacked someone with a weapon*

is susceptible to change

DMD Results

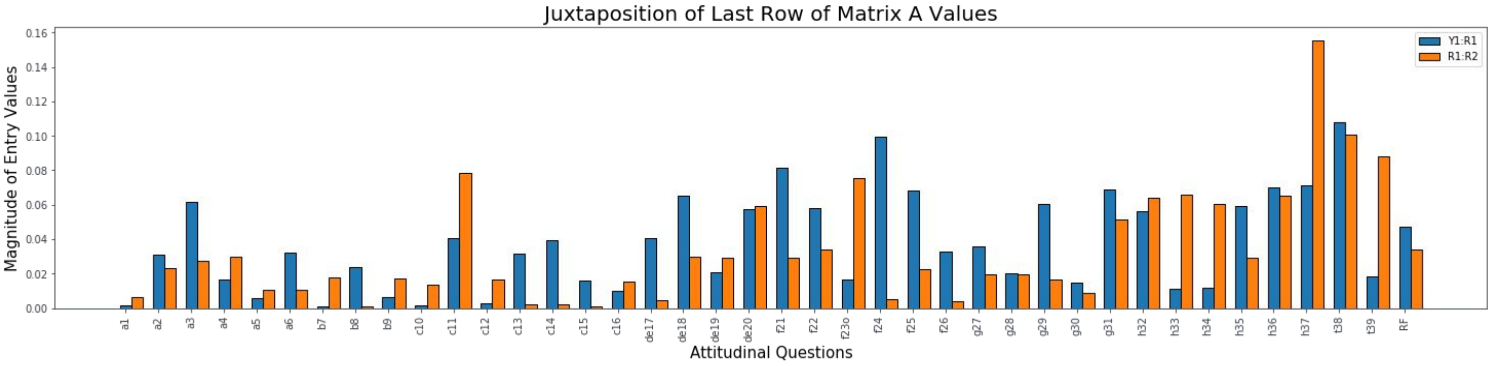
Questions’ Contributions to Risk Factor

Purpose: Analyze **importance** of questions in calculating the total risk score of the next program period

$$X_0, X_1 = \left[\begin{array}{c} QuestionResponses \\ \\ \hline RiskFactor \end{array}\right]$$

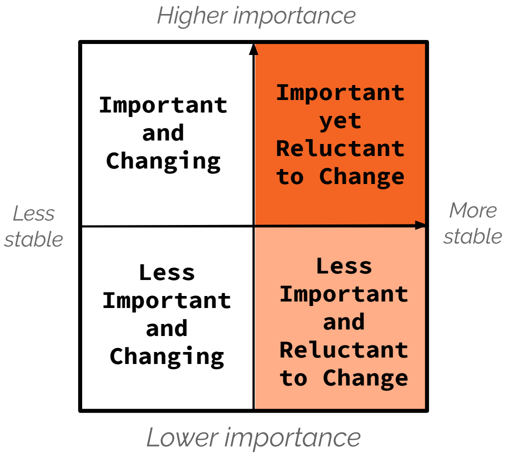
$$AX_0 = X_1 \qquad (1)$$

$$\left[\begin{array}{c} \\ \hline \boxed{LastRow} \end{array}\right] \left[\begin{array}{c} \\ \hline Risk \end{array}\right] = \left[\begin{array}{c} \\ \hline \boxed{}\ FutureRisk \end{array}\right]$$

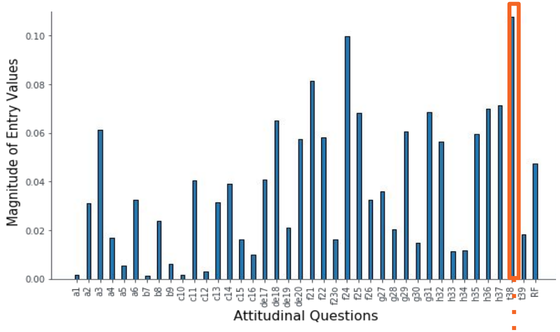


DMD Results

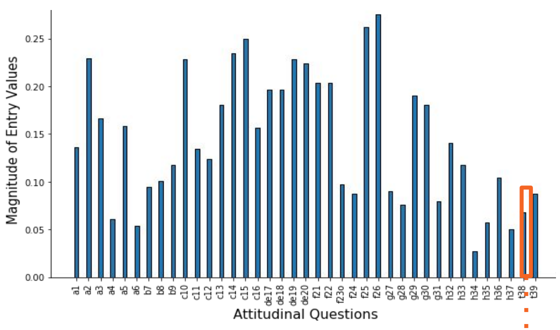
Change vs. Importance

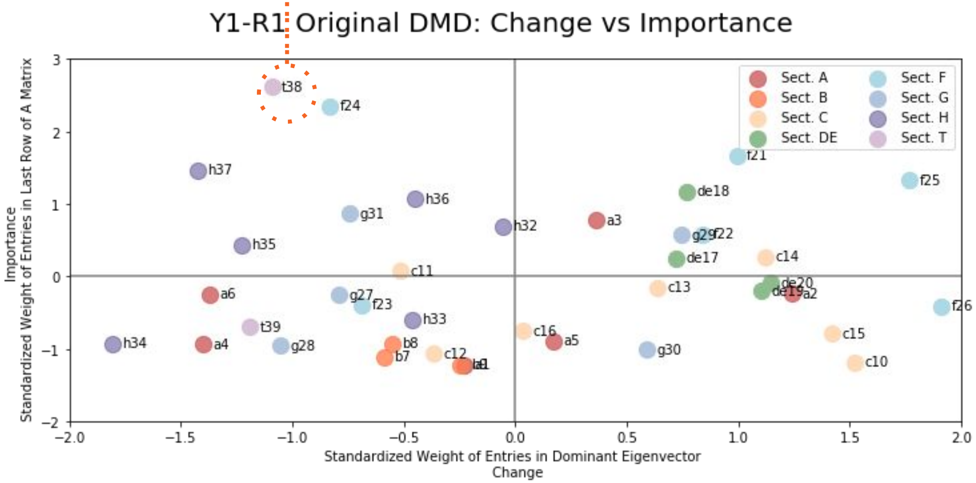


Change: dominant eigenvector entries



Importance: last row entries of matrix A

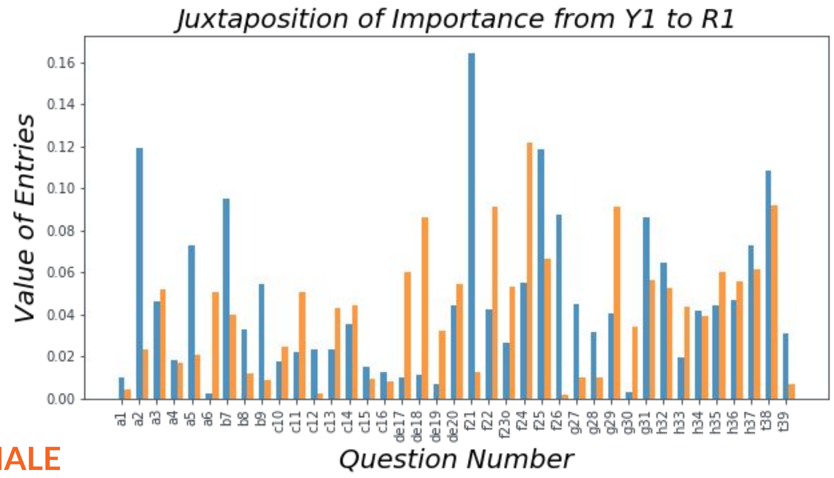


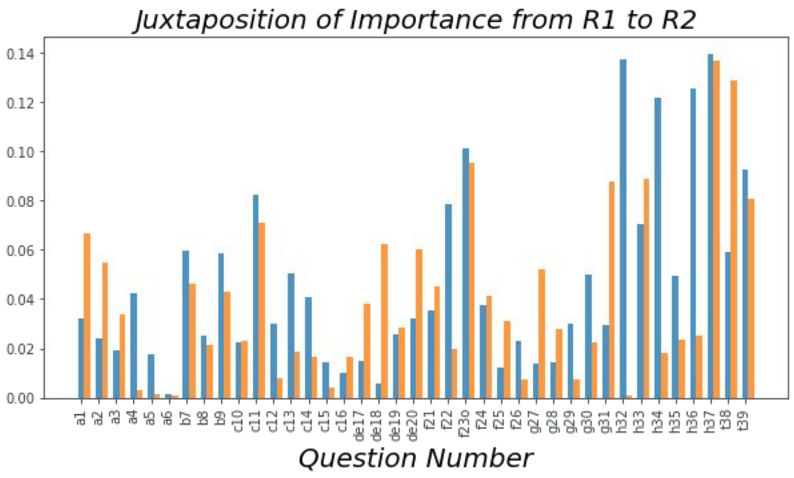


Shortcomings of DMD

The DMD algorithm does not explain the **variation** in change among different **characteristics** of participants like age and gender, and this issue can be resolved by applying DMD with control (DMDc)

Gender Differences



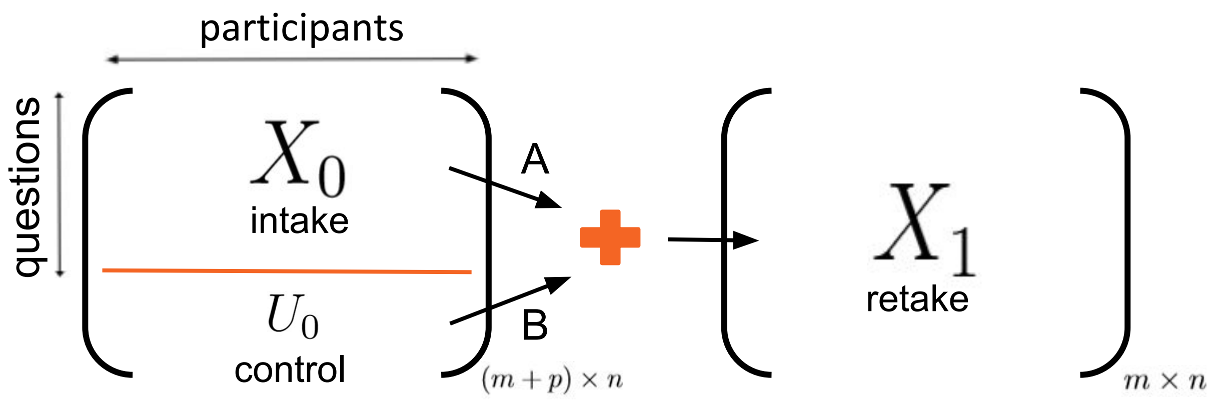


FEMALE
MALE

Dynamic Mode Decomposition + Control

DMDc

Algorithm



m = number of questions; n = number of participants; p = number of control levels

1) Find transformation matrices **A** and **B** such that **($AX_0 + Bu_0$)** approximates **X_1**: $\| AX_0 + Bu_0 - X_1 \|_F \simeq 0$

2) Compute Singular Value Decomposition of **X_0**: **$X_0 = U\Sigma V^*$**

3) Then compute for transition matrix **$A = X_1 V \Sigma^{-1} U_A^*$**, and **$B = X_1 V \Sigma^{-1} U_B^*$**, where **$U_A$** is the first m rows of **U**, and **U_B** is the last p rows of **U**

4) Make predictions using **$X_1 = AX_0 + Bu_0$**

Control Variables

Gender

- Male
- Female

Ethnicity

- African-American
- Asian
- Hispanic
- White
- Other

District

- 23 GRYD Zones

Age

- 10-16 years old

Control matrices created using one-hot encoding:

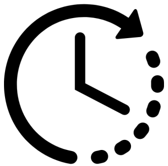
$$\begin{array}{c} \begin{array}{cc}\text{Male} & \text{Female}\end{array} \\ \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\end{array}$$

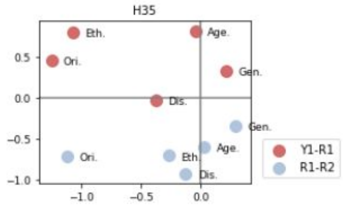
E.g. (for gender)

DMD & DMDc

Change vs. Importance

What we have learned...





> Questions' **importance** typically **change over time**

> Different questions have different effects according to **demographic groups**

> Effects are shown in results of models with different **controls**

Important and Changing	**Important yet Reluctant to Change**
Less Important and Changing	**Less Important and Reluctant to Change**

For all participants, GRYD should put more emphasis in areas of:

Section C
Critical Life Events

Section DE
Impulsive Risk Taking

Section H
Peer Delinquency

Section T
Family Gang Influence

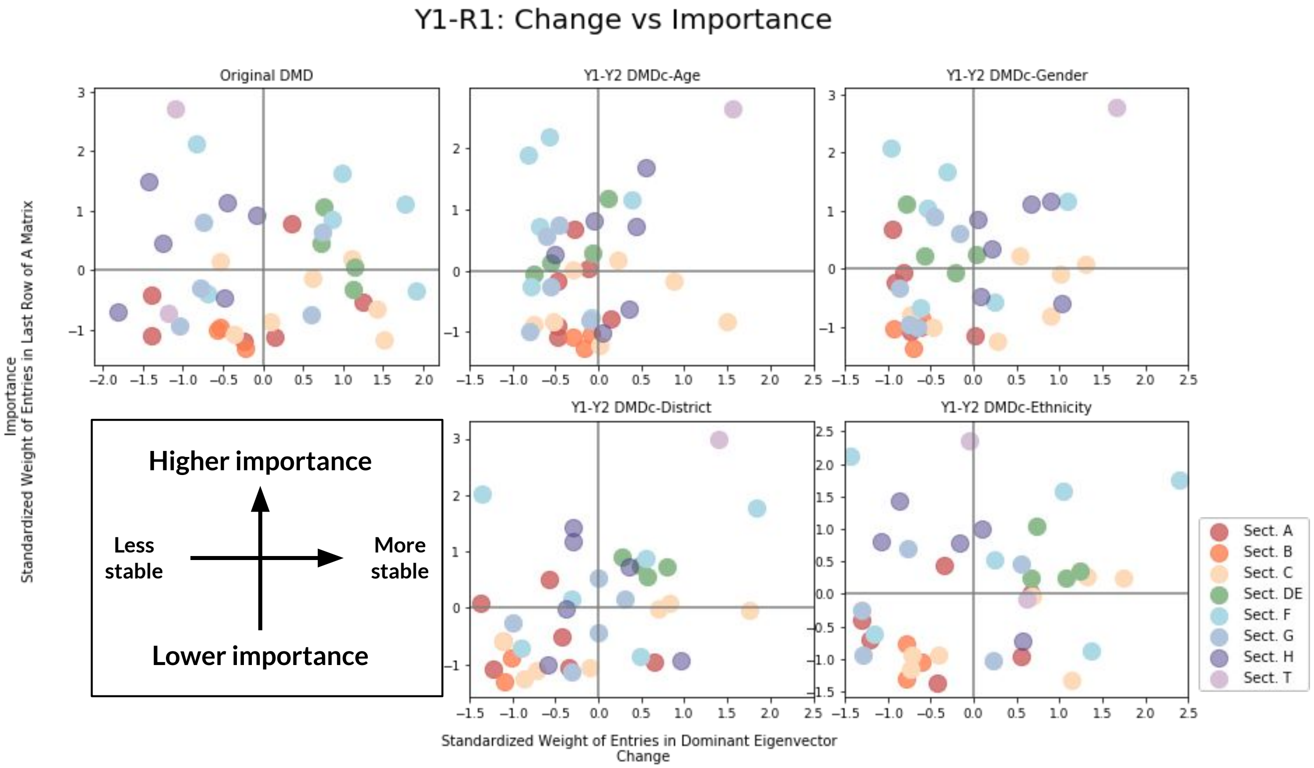
DMD & DMDc

Change

vs.

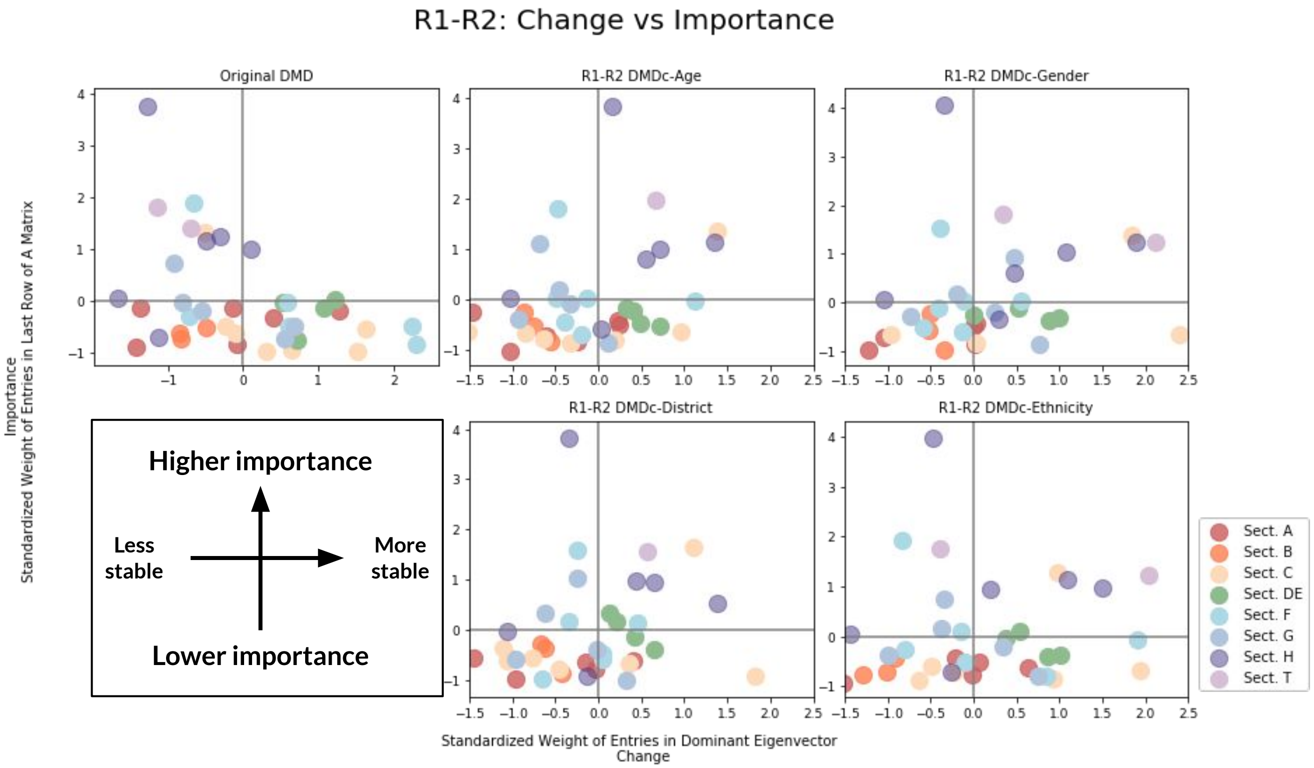
Importance

Important and Changing	Important yet Reluctant to Change
Less Important and Changing	Less Important and Reluctant to Change



Change
vs.
Importance

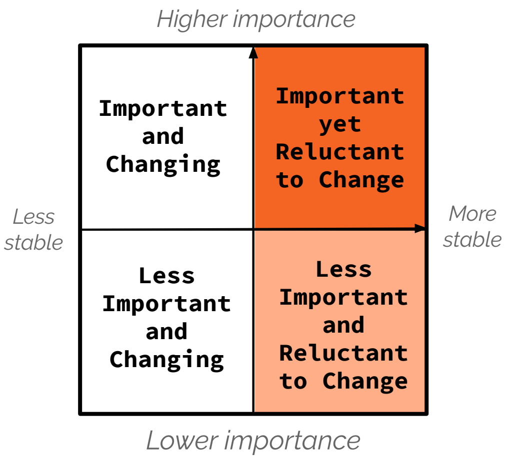
Important and Changing	Important yet Reluctant to Change
Less Important and Changing	Less Important and Reluctant to Change

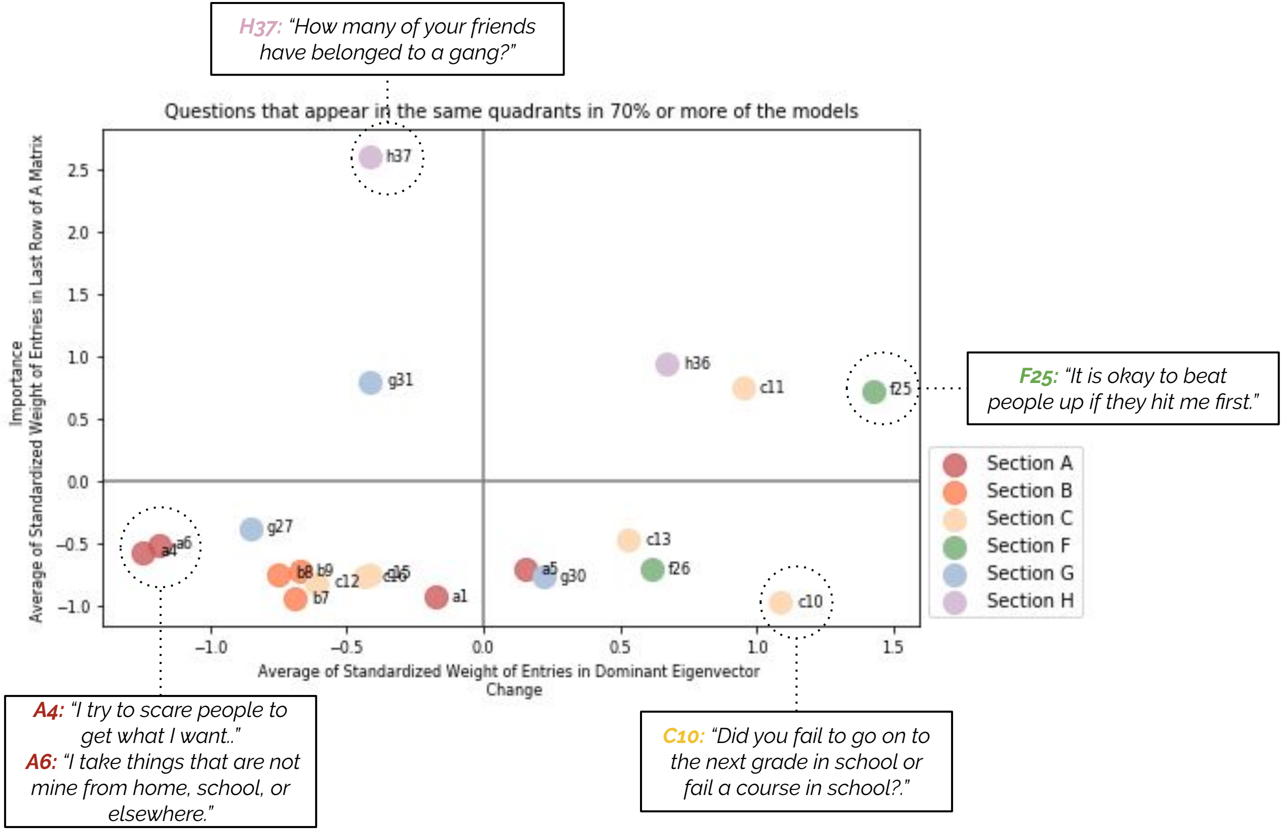


DMD & DMDc

Change
vs.
Importance

Questions that consistently appear in the **same quadrants** across models

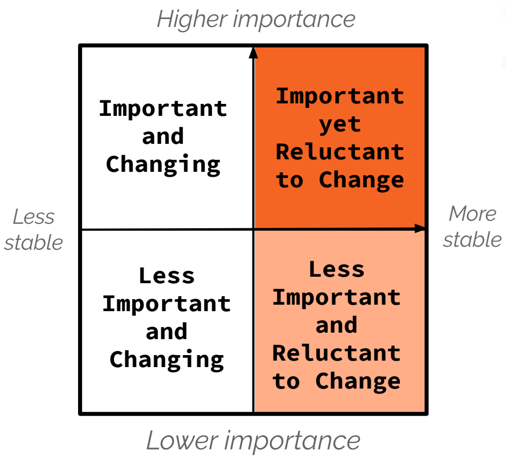


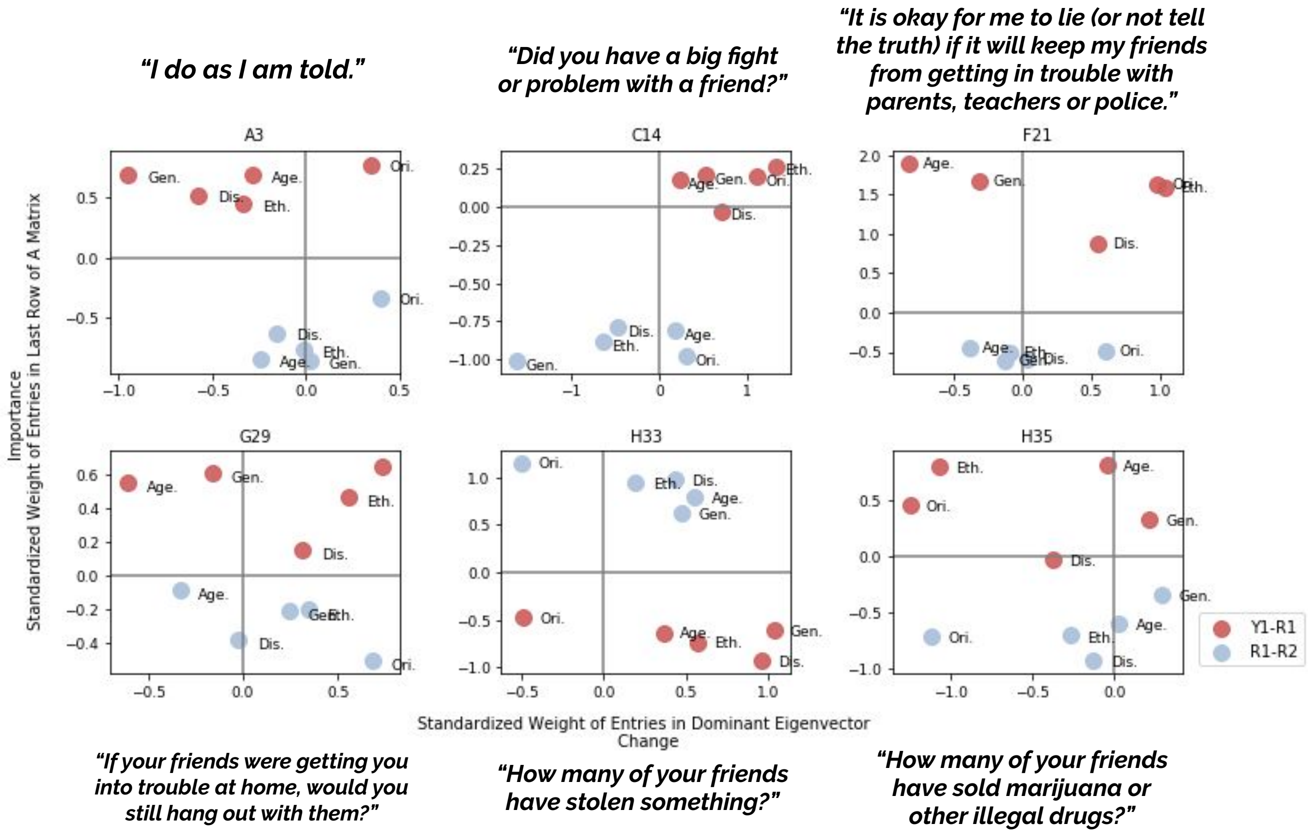


DMD & DMDc

Change
vs.
Importance

Questions that appear in **different quadrants** in different models

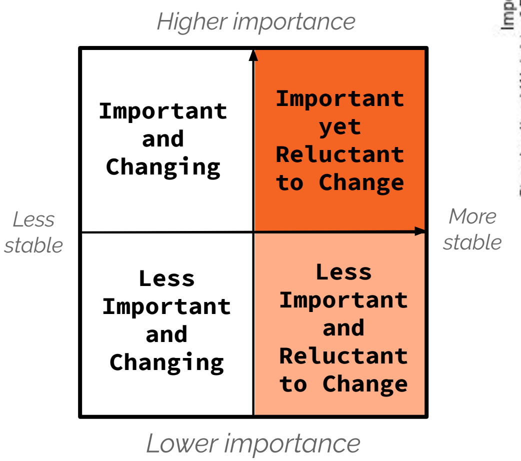


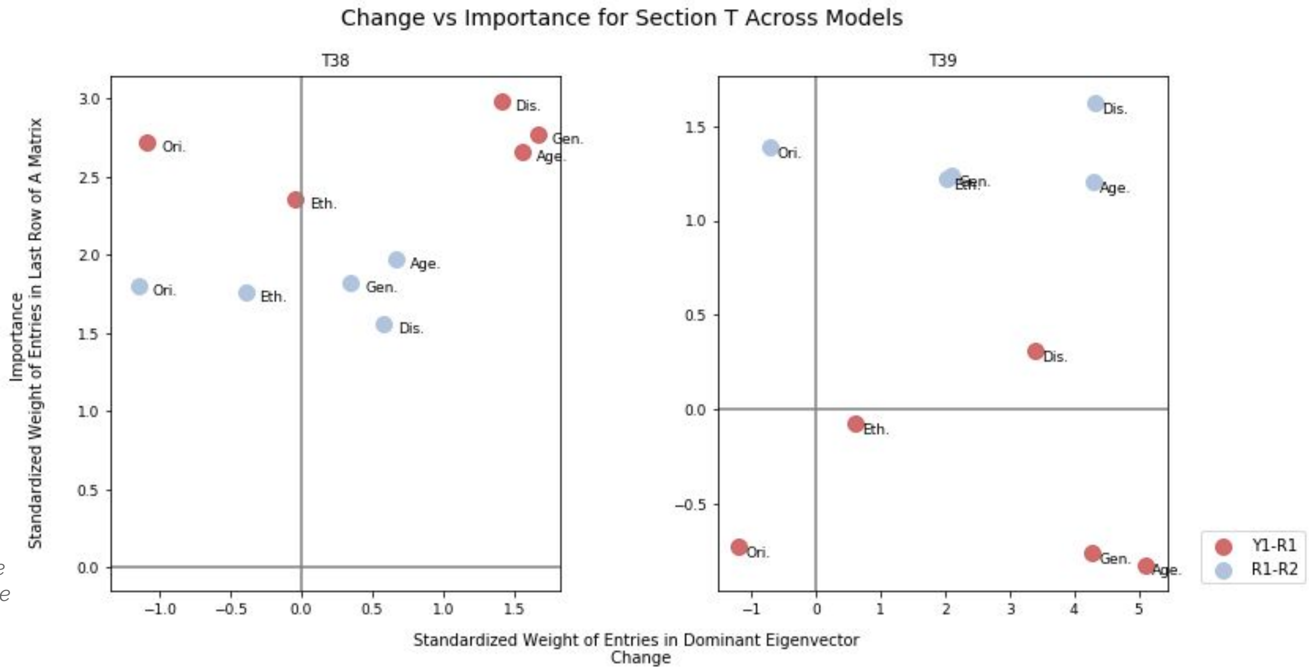


DMD & DMDc

Change
vs.
Importance

Section T
Family Gang Influence





"How many people in your family think that you will join a gang?"

" How many people in your family are gang members?"

Error Minimization Properties

DMD

Lemma 1: Let X be a $k \times n$ matrix, and Y another $m \times n$ matrix, then the matrix $A \in \mathbb{C}^{m \times k}$ that minimizes the functional $g(A) = \|AX - Y\|_2^2$ is given by the solution of the normal equations $AXX^* = YX^*$.

Lemma 2: $A = YX^\dagger$ is a solution to the equation $AXX^* = YX^*$.

DMDc

Lemma 3: Let X_{k+1} a real, $m \times n$ matrix, X_k another real $j \times n$ matrix, and u_k a $k \times n$,matrix be given, let Y to be the $(j+k) \times n$ matrix by appending u_k to X_k, then there exists some $M \in \mathbb{C}^{m \times (j+k)}$ the solution to $MYY^* = X_{k+1}Y^*$ of the matrix $A \in \mathbb{C}^{m \times j}$ and the matrix $B \in \mathbb{C}^{m \times k}$ that minimizes the functional $g(x) = \|X_{k+1} - AX_k - Bu_k\|_2^2$ are given by the first j columns of M and last k columns of M, respectively.

— DMD & DMDc

Prediction Accuracy

Randomly split data into 80% training and 20% testing

Run t-test against DMD to demonstrate the significant increase in precision.

DMDc on district has the smallest MSE

DMD (No Control)		
Trial	MSE (Test)	MSE (Whole)
1	0.1018	0.0965
2	0.1004	0.0964
3	0.1012	0.0964
4	0.1006	0.0964
5	0.0993	0.0964

DMDc (Age)		
Trial	MSE (Test)	MSE (Whole)
1	0.0981	0.0948
2	0.0983	0.0949
3	0.1025	0.0949
4	0.1009	0.0950
5	0.1020	0.0950

DMDc (District)		
Trial	MSE (Test)	MSE (Whole)
1	0.0988	0.0924
2	0.1009	0.0924
3	0.1007	0.0924
4	0.1028	0.0923
5	0.1019	0.0924

DMDc (Ethnicity)		
Trial	MSE (Test)	MSE (Whole)
1	0.0990	0.0956
2	0.1030	0.0956
3	0.1028	0.0957
4	0.0995	0.0957
5	0.1025	0.0957

DMDc (Gender)		
Trial	MSE (Test)	MSE (Whole)
1	0.1027	0.0957
2	0.1014	0.0956
3	0.1007	0.0956
4	0.0999	0.0956
5	0.0988	0.0956

Table 2. MSE of DMD and DMDc trials

DMD and DMDc (Age)		
MSE	t statistic	p value
Test	4.1337	0.0001
Whole	232.4665	0.0000

DMD and DMDc (District)		
MSE	t statistic	p value
Test	6.9643	0.0000
Whole	403.1224	0.0000

DMD and DMDc (Ethnicity)		
MSE	t statistic	p value
test	0.4925	0.6229
Whole	121.0410	0.0000

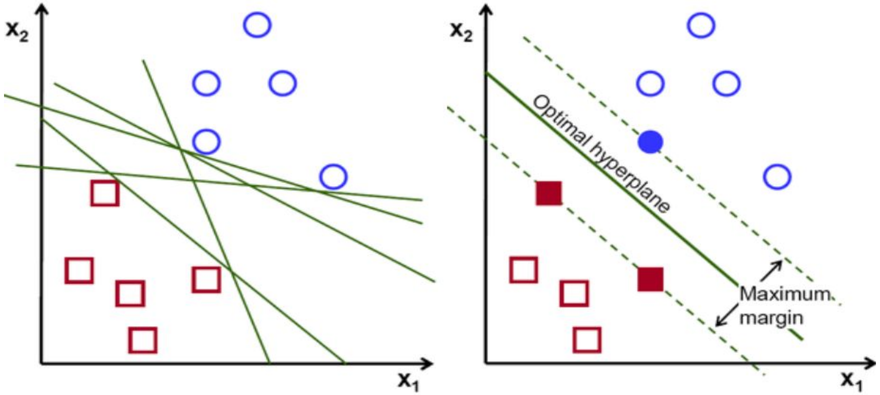
DMD and DMDc (Gender)		
MSE	t statistic	p value
Test	4.4854	0.0000
Whole	148.0830	0.0000

Table 3. t-test of DMD and DMDc

Confirmation of DMD via Machine Learning

using SVM + DT

Support Vector Machine (SVM)

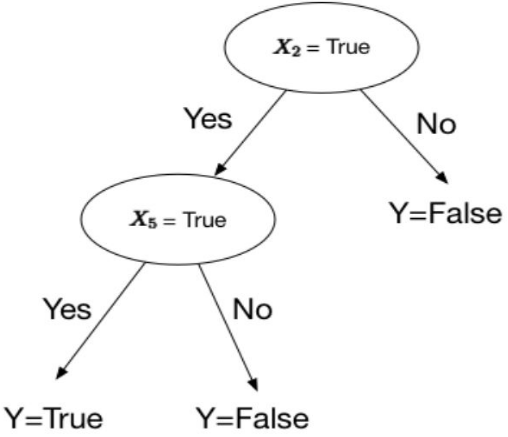


$$\min_{\mathbf{w},\xi,\rho} \frac{1}{2}\mathbf{w}^T\mathbf{w} + \sum_{i=1}^{n} C_i\xi_i$$

$$\text{subject to } y_i(\mathbf{w}^T\phi(\mathbf{x}_i)+\rho) \geq 1-\xi_i, \quad i=1,\ldots,n$$

$$\xi_i \geq 0, \quad i=1,\ldots,n$$

Decision



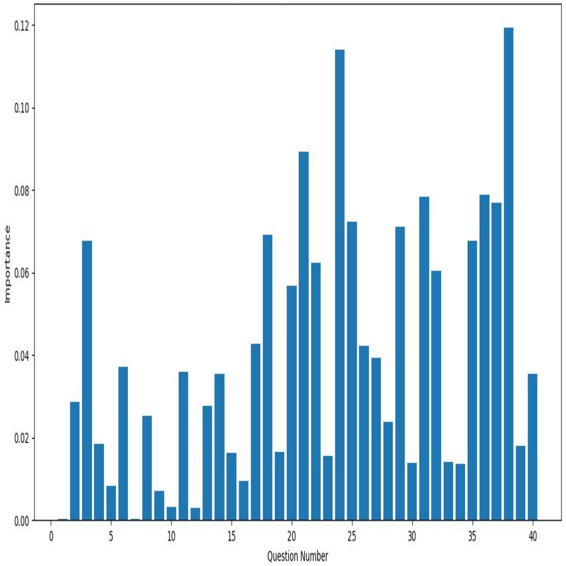
Key Concepts:

- Entropy Loss
- $E(S) = \sum_{i=1}^{c} -p_i \log_2 p_i$

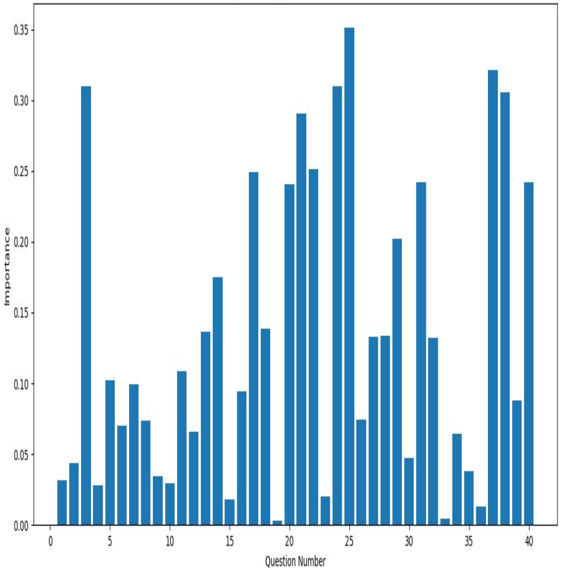
Confirmation of Question Importance

DMD vs SVM

DMD



SVM



Statistical Evidence
Pearson Correlation
0.7323
Spearman Correlation
0.6548
Test of Spearman Correlation
0.000

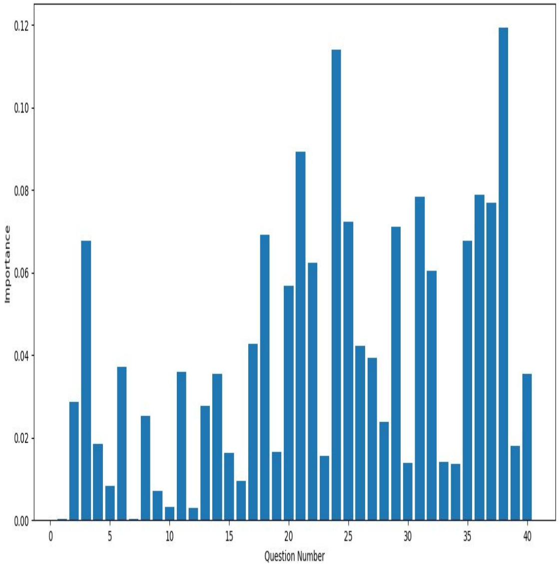
Question Importance from DMD and SVM

Null hypothesis:
results for the two models are uncorrelated

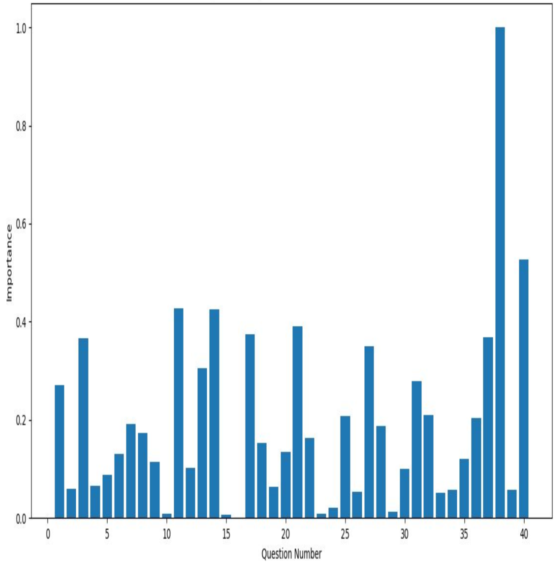
Confirmation of Question Importance

DMD vs DT

DMD



DT



Statistical Evidence
Pearson Correlation
0.4675
Spearman Correlation
0.4315
Test of Spearman Correlation
0.005

Question Importance from DMD and DT

Null hypothesis:
results for the two models are uncorrelated

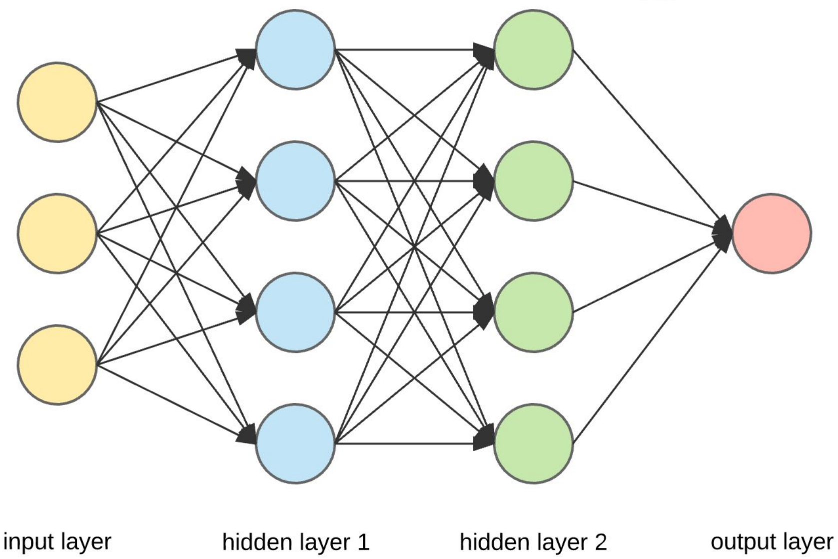
Future Work

1. Explore the relationship between DMD algorithm and neural networks
2. Find a criterion for detecting the dropout group from the program
3. Try develop a modified version of DMD which take partial label information

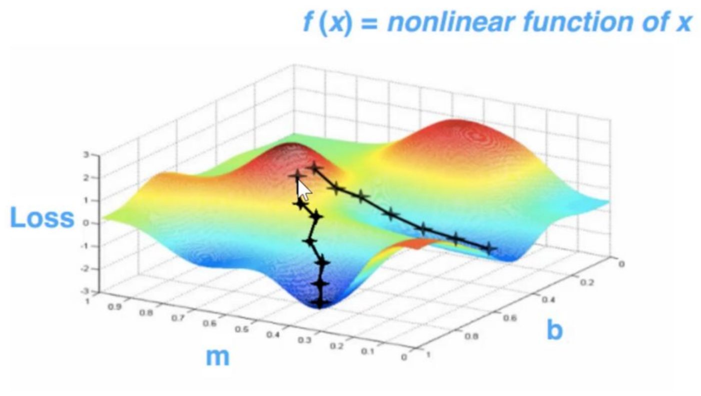
— Future Work

ANN and Gradient Descent

$$h = w \cdot x + b \qquad o = f(h)$$



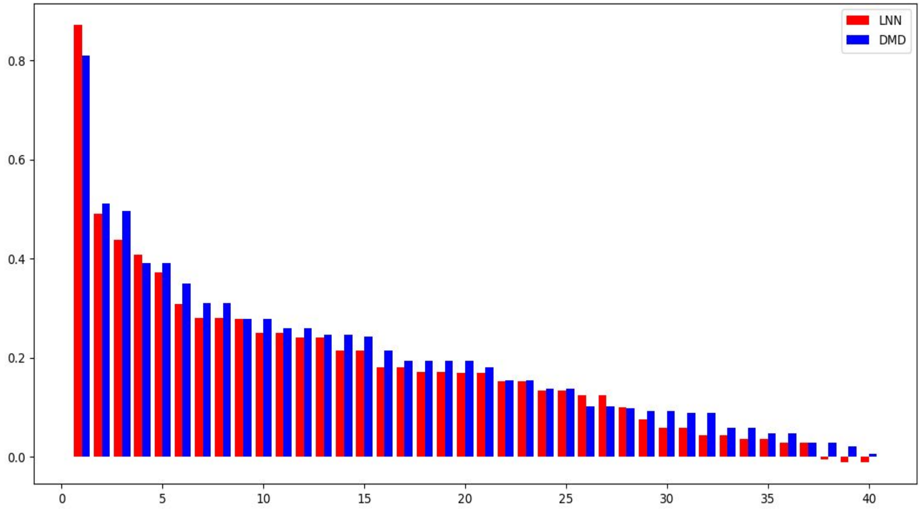
Basic Structure of **Artificial Neural Network**



Adam:
An algorithm that is widely used in stochastic optimization problem in Neural Network
AdaGrad:
An adaptive gradient method used in optimization problem

— Future Work

Comparison of DMD and LNN



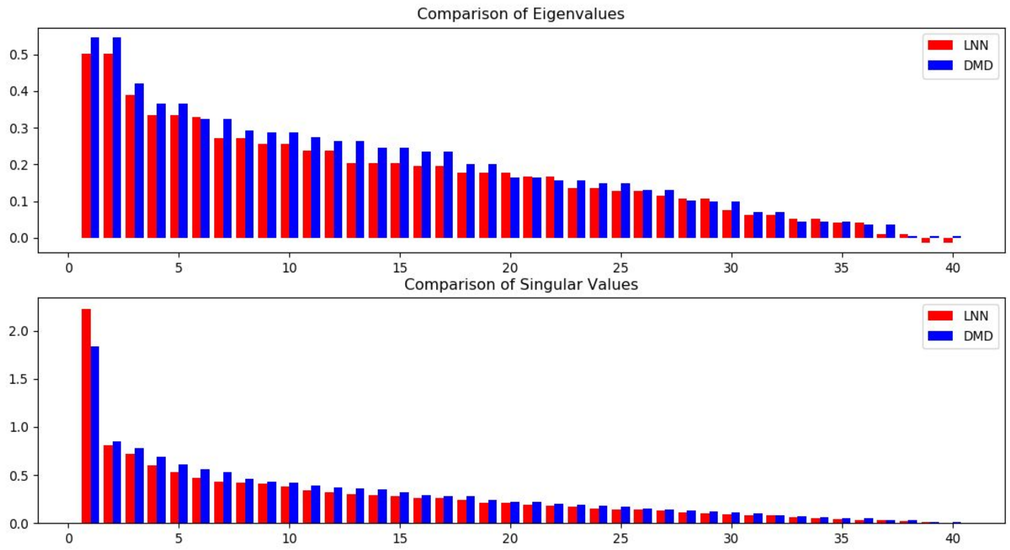
	DMD		LNN	
	Test	Whole	Test	Whole
Trial 1	0.1007	0.0965	0.1000	0.0966
Trial 2	0.1010	0.0964	0.1007	0.0968
Trial 3	0.1004	0.0964	0.1011	0.0967
Mean	0.1013	0.0964	0.1018	0.0975
S.D.	0.0017	0.0004	0.0020	0.0005

Comparison of Eigenvalues
Correlation Coefficient = 0.9988

Comparison of MSE

Future Work

Comparison of DMDc and LNNc



	DMDc		LNNc	
	Test	Whole	Test	Whole
Trial 1	0.0987	0.0956	0.1004	0.0962
Trial 2	0.0996	0.0956	0.1008	0.0965
Trial 3	0.0984	0.0956	0.1011	0.0972
Mean	0.1010	0.0956	0.1016	0.0968
S.D.	0.0018	0.0005	0.0020	0.0008

Comparison of Eigenvalues and Singular Values
Correlation Coefficient (Eigenvalues) = 0.9927
Correlation Coefficient (Singular Values) = 0.9864

Comparison of MSE

— Future Work

Discussion of the Theoretical Lowest Error Bound

Objective Function of DMD: $X_{k+1} = AX_k$

Objective Function of LNN(c): $X_{k+1} = AX_k + b$

Objective Function of DMDc: $X_{k+1} = AX_k + Bu_k$

Optimization Problem of DMD: $L_1 = \min_{A} ||x_{k+1} - Ax_k||_2$

Optimization Problem of LNN(c): $L_2 = \min_{A,b} ||x_{k+1} - Ax_k - b||_2$

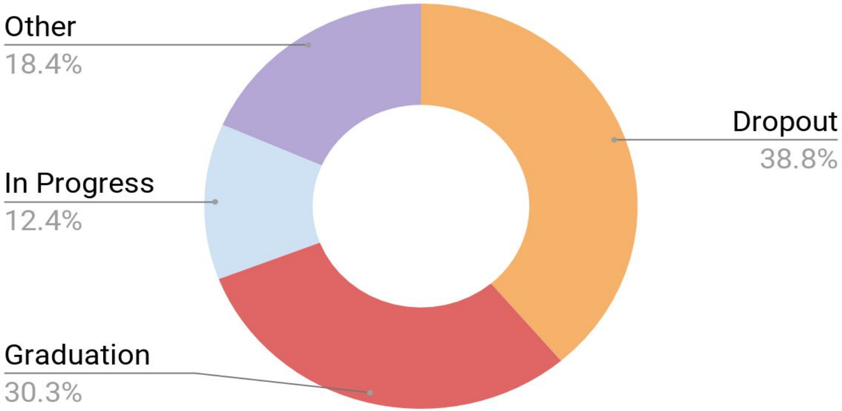
Optimization Problem of DMDc: $L_3 = \min_{A,B} ||x_{k+1} - Ax_k - Bu_k||_2$

Relationship of Lowest Error Bound of Three Methods: $L1 \geq L2 \geq L3$

— Future Work

Distinguishing Dropouts

Distribution of the Status



Summary of Program Dropout	
Number of Participants	22548
Participants with Status	13549
Number of Dropouts	5258
Dropouts after Y1	4110
Dropouts after R1	922
Dropouts after R2	226

Possible Reasons

Long Term Non-Attendance

Refuse Service

	SVM	Random Forest
Acc.	85.89%	89.42%

Future Direction:
- Interpretable rules of how to determine whether participant will dropout
- Metrics describing the probability of dropout

Recap

DMD

- Used to observe change in question responses & their contribution in calculating risk factor

DMDc

- Used to make these same observations as DMD, however includes <u>control</u> variable
- Has a <u>higher prediction accuracy</u> than DMD

Confirmed DMD/DMDc algorithm via machine learning

- SVM and Decision Tree provide <u>correlated</u> results of question importances

*Question scores are **decaying** over time.*	*GRYD should focus on questions that are **important, but reluctant to change**.* *(Critical Life Events, Impulsive Risk Taking, Peer Delinquency, and Family Gang Influence)*	*Also we need to take into account **demographic groups** because they do **have an effect** on question responses.*

Thanks for listening!

Special Thanks to:

Professor Jeffrey Brantingham and Professor Andrea Bertozzi

Questions?