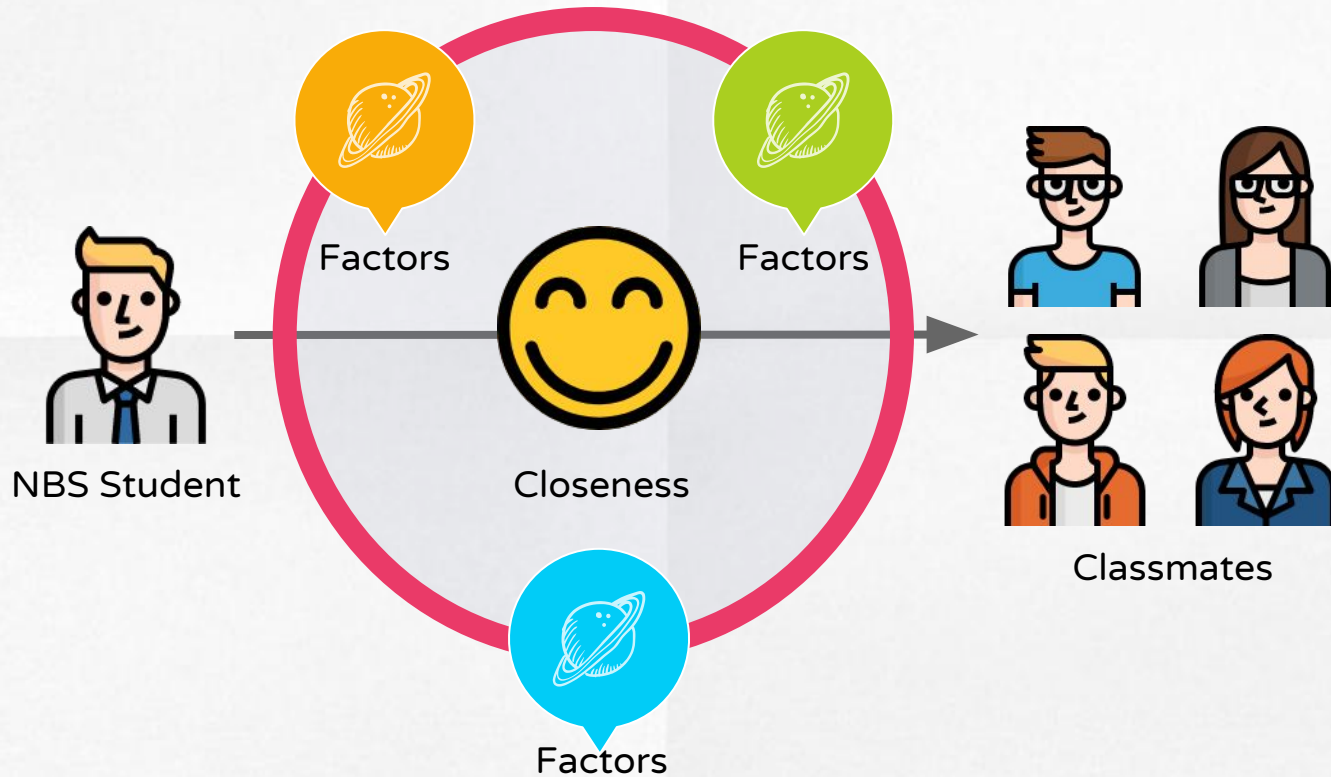


How Close Are NBS Students To Their Classmates?

Cheng Jin Yee (Jinny)
Jeremy Jerome Chia

CE9010 Introduction to Data Science
AY1718 S2

I. Data Problem



2. Data Acquisition

Friends in NBS

This survey is targeted at students in NBS - and aims to examine the avenues in which friendships are formed and sustained throughout one's course of study in NBS.

The results of the survey will be used to influence the design of programmes for this coming Freshmen Orientation Programme. Your honest feedback will be used to make decisions based on our experiences. Your honest feedback will be used to make decisions based on our experiences. This form does not collect your e-mail.

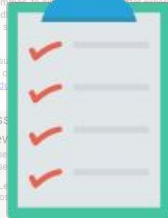
If you are interested in the results of the survey, please email me at jchia019@e.ntu.edu.sg and I would gladly mail you a copy of the results. If you have any queries, please email me at jchia019@e.ntu.edu.sg.

Rate your closeness to the following people on a scale of 1-10 based on how often you meet them. (e.g. if you're close to someone, rate them 10; if you're not close, rate them 1.)

Relationship	1 (Least Close)	2	3	4	5	6	7	8	9	10 (Most Close)
Classmates (from Index)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Faculty	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Camp Mates	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Big 4 Camp Mates	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hall	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
CCAs	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

What makes me feel close to my friends? *

Your answer



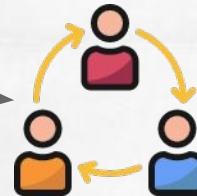
Survey

"Friends in NBS"

Factors



How Often Do You Meet?

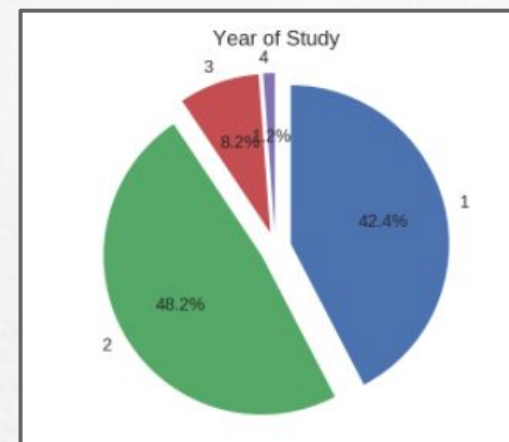
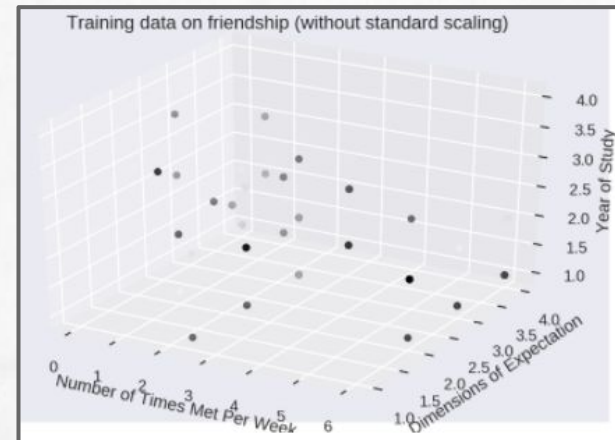
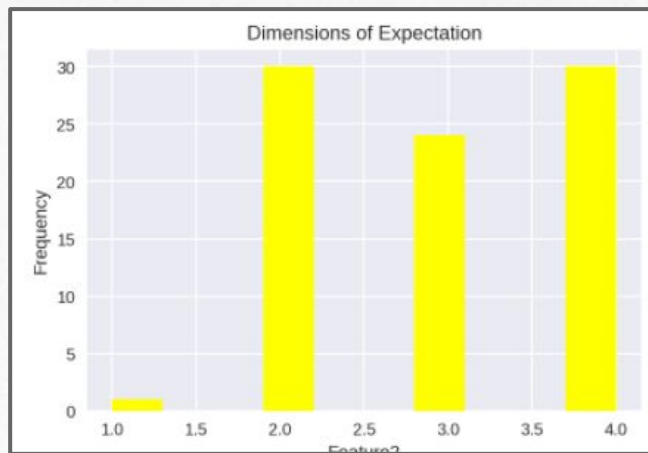
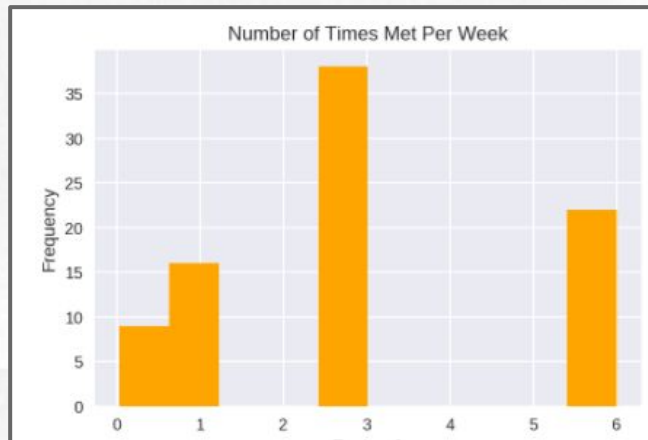


Dimension of Expectations

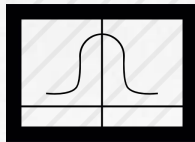


Year of Study

3. Data Exploration

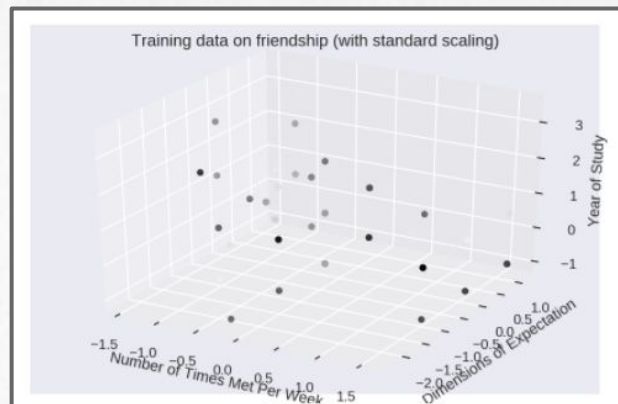


4. Pre-Processing



z-Scoring
of Data

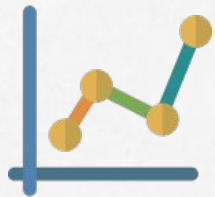
$$z_i = \frac{(x_i - \bar{x}_i)}{\sigma_i}$$



Removal of
Noise Data

What makes me feel close to my friend	How often do I see my friend	Let my friend know my feelings	Can my friend help me
Common interest/topic	Every week		6
	2-3 Times a Month		8
same timetable	2-3 Times a Week		8
Our friendship	Every week	kkwua	10
their sincerity	2-3 Times a Week		10
Things done	2-3 Times a Week		3
It makes me feel close to my friends when	Almost every day	yloke	10
More time together	Almost every day		9
Frequent interaction	2-3 Times a Week	ctay0	10
They cares about me	Almost every day		6
Time spent hanging out, common talking to	2-3 Times a Week		3
depth	2-3 Times a Month		7

5. Data Analysis

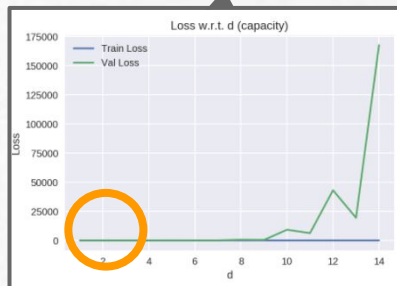


Regression Analysis

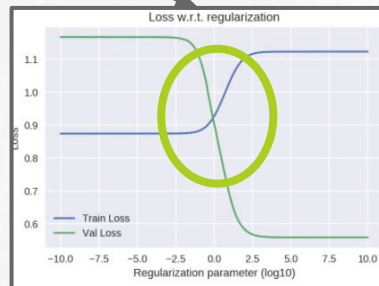
$$y_i = w_1 x_1 + w_2 x_2 + w_3 x_3 + \dots$$

Evaluated By

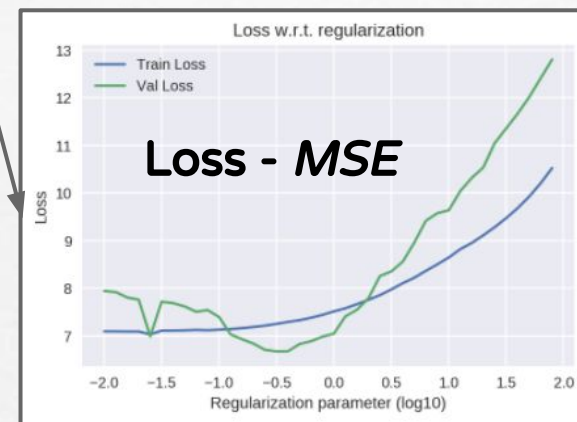
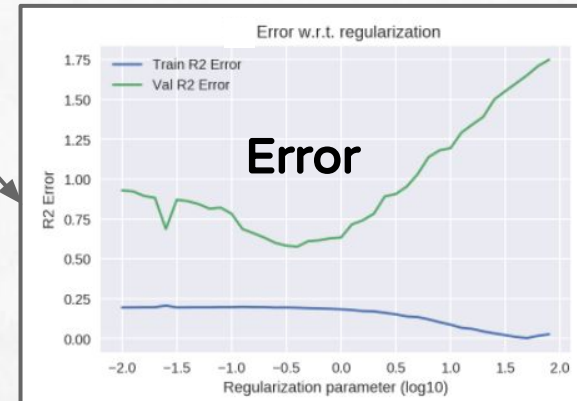
Techniques Used



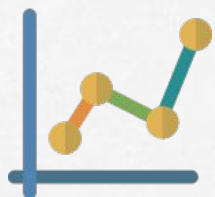
Overfit training data and finding balance



Add regularisation to the validation set and evaluate generalisation of the validation set



5. Data Analysis



Regression Analysis

$$y_i = w_1x_1 + w_2x_2 + w_3x_3 + \dots$$

Variations

- Without Normalisation
- Reducing Numbers of Features
- K-Fold Cross Validation

Technique	Optimal d	λ	Train Loss	Val Loss
With Normalisation	2	0.794	0.91	0.92
Pre-processing	2	0.316	7.25	6.67
Removal of Feature 1	3	0.398	1.00	1.01
Removal of Feature 2	2	0.316	0.94	0.94
Removal of Feature 3	2	1.000	0.98	0.68
K-Fold Cross Validation	1	0.100	0.87	0.85

**BEST
MODEL!**

6. Analysis of Results



Observations

- **Normalisation** increases the performance of the model.
- **Regularisation** strikes a balance between bias and variance, and generalises the well-fitted model.
- **Cross-Validation** reduces the loss.
- **Data Features** may not be the most suitable, though there is evidence of some relationship.
 - Loss does not go below 0.6 for the training data even after overfitting; may not be the most appropriate features.



Improvements

- **Collect More Data**
 - Reach maximum learning capacity
 - At high capacity, collecting a large number of data will not cause overfitting and can greatly reduce loss (fix high variance)
- **Select more representative features**
 - Handcrafting Features
 - Fix High Variance

The END