Yutian Lai Assignment1

Building models

Deviant aggressive behavior

1. What social policy would be appropriate to reduce deviant aggressive behavior if Theory I were correct? Theory II? Theory III? Theory IV?

If theory I were correct, reward and punishment system should be taken advantage of to reduce deviant aggressive behavior because this system is important in managing the rational operation of a society. Social policies should on one hand, reward good behaviors and on the other hand, penalize deviant aggressive behaviors. Under such policy, people would learn from experience that they could gain benefits when behaving good, such as stopping others' bullying behavior, thus getting motivated to repeat the good action, or they would receive penalty when conducting deviant aggressive behavior, such as to end up with death penalty if they fought someone to death, so they would avoid such deviant aggressive behavior. Education should also be strengthened to teach children which behaviors are good and which should be avoided by receiving timely feedback from their teachers, so as to decrease the possibility that they would conduct deviant aggressive behaviors when they grow up.

If theory II were correct, deviant aggressive behavior is the result of people's hostility and frustration against the authorities over them. Social policies should aim at helping people find reasonable ways to alleviate their stress and frustration, to bridge the gap and build a more harmonious relationship between frustrated individuals and authority figures. For instance, At company level, authorities should express care for their employees, provide them with free counselling services that help reduce their frustrations and improve their mental health status. In this case, the employees' hostility towards their employers would decline or even disappear. At family level, parents should enforce communication with their children, to spend a certain amount of time each week listening to their children's thought and to lower down the pressures put upon them. In this way, the tension between frustrated ones and authorities would be eased and deviant aggressive behavior would be less.

If theory III were correct, deviant aggressive behavior is the reasonable response of systematically discriminated people. Social policies should aim at eliminating social discrimination. Since the situation cannot be changed instantly owing to the complicated interweave of different factors contributing to discrimination, such as race, gender, economic status, etc.. The government should carry out both short-term and long-term policies. In the short term , policies should focus more on compensating for the currently discriminated ones. For instance, the government should grant children living in slums enough financial welfare and opportunity of

education, and then chances would decrease that they would conduct deviant behavior in the future. In long term, the focus should switch to “equality” , to refrain and finally erase systematic discrimination. For instance, the government could give companies certain benefits, such as extra tax allowance if they hire more people in the discriminated gender or racial groups; schools should also teach children not to discriminate others and the importance of “ equality” since their early age. When the society becomes fairer and the discriminated groups gain the same education and employment opportunities as others, they would start to benefit from behaving rationally and following social rules, so the deviant aggressive behavior would be mitigated.

Finally, if theory IV were correct, deviant aggressive behavior is a social role which individuals in contact with deviant subcultures would get involved into. Social policies aiming at restricting the influence of deviant subcultures and further preventing the formation of such subcultures would be effective. For restriction, access to websites propagating deviant subcultures should be banned and social organizations promoting deviant cultures should be made illegal. For prevention, if the government can try best to fulfill people’s legitimate needs, they might not be too dissatisfied or frustrated to form such subculture to damage the rational operation of society. The government can also take effort to build a more positive cultural environment fitting mainstream social norm, so as to reduce the possibility that people deviate into undesired small subculture groups. In this way, without contact with deviant subcultures, individuals would not be socialized into the deviant role and deviant aggressive behaviors would dissipate.

Waiting until the last minute

(a) Ask yourself why the observation might be true and write down your explanations.

Subjective reasons for procrastination might include lack of self-confidence, seek for the feeling of accomplishment, pursuit of efficiency, dislike for the assignment and low self-expectation. First, people might start to do things at the last moment because they do not believe they can complete the task well. For instance, people may put off writing a paper because they do not believe they can get a good grade from their work, or putting off doing a problem set because it is highly likely they would not get the answers right. Second, to start late and finish the assignment in a rather short period would bring people extra pleasure as they have finished a challenging and risky task. In this way, they would gain more feelings of self-accomplishment and self-competence. Third, procrastination might stem from the pursuit of efficiency. Conventionally, people have the propensity to be more efficient when approaching the deadline owing to the fear of not finishing work on time. So it would take less time to finish a task if one starts at a late stage. Moreover, People may also refuse to start the task because they dislike it and would like to spend as less time

in doing things unenjoyable as possible. Finally, procrastination behavior might be correlated with one's low self-expectation. For instance, if one shoots for a "Pass" rather than "A" for a particular task, he/she would start late because he/she doesn't really care about the result.

Objective reasons for procrastination, on the other hand, emerge from the characters of the tasks themselves. "Last-minute" tasks are usually long-term, having high proportion in final score, inspiration-needed, highly difficult, or combined with many other tasks at the same time. First, if an assignment is released a long period before deadline, people might procrastinate because they believe they would always have time to do the task in the future, or lack of patience to wait for the grades/results for a long time if they finished the task too early. Second, if the task is worthy of high points, people might start late because they would take the task seriously and would make best preparation out of the pursuit of perfection, worried about screwing it up if starting early. Meanwhile, starting early doesn't necessarily mean doing the assignment well, as some assignments, such as writing papers, require inspiration. And the moment inspiration hits might be right before the deadline. Another concern is the assignment's degree of difficulty, as high difficulty might push people to start at the last minute because they might not know where to start, or never start until getting help from the office hour, or struggle to gain extension from the instructor rather than following the deadline. In addition to these factors, another possible explanation is the number of other assignments one have at the same time. If one has to finish multiple tasks at the same time, it is highly likely the assignment would be completed right before the deadline.

(b) Generalize the explanatory model – that is, induce the most general, abstract model you can produce that still has the original observation as a consequence.

The model below is built to examine whether people postpone doing a task owing to the five **subjective** explanatory variables I listed in (a):

x_1 : self-confidence ; x_2 : willing to seek feeling of accomplishment

x_3 :pursuit of efficiency; x_4 :dislike for the particular assignment

x_5 : self-expectation

I'm going to use a scale from 1-5 to measure the five variables, with the question being "Which degree of these factors do you think you have?", and the five scales respectively correspond with "very low", "low", "medium", "high" and "very high".

To make the measurement more consistent, for variables x_1 and x_5, I'm going to

start from 1 as “very high” to 5 as “very low”. So as my explanation made in (a), degree of procrastination should increase as these index increase.

The outcome variable, Y, measures how early people start working on a task by subtracting starting time from deadline time. The lower the outcome variable, the higher degree of procrastination one has.

The model is shown below, with β_0 being y-intercept (constant term), β_i being slope coefficients for each explanatory variable and ε being the model’s error term.

$$Y=\beta_0 +\beta_1 x_1 +\beta_2 x_2 +\beta_3 x_3 +\beta_4 x_4 +\beta_5 x_5 +\varepsilon$$

(c) Induce an alternative model that also has the original observation as a consequence.

The model below is built to examine whether people postpone doing a task owing to the five **objective** explanatory variables I listed in (a):

x_1: the duration between the time the assignment is released and the deadline

x_2: proportion of the assignment in final grades

x_3:how much the assignment requires inspiration; x_4: degree of difficulty

x_5: loads of other assignments with similar deadline

I’m going to use a scale from 1-5 to measure the five variables, with the question being “Which degree of these factors does the assignment have?”, and the five scales respectively correspond with “very low(very short)”, “low(short)”, “medium”, “high(long)” and “very high(very long)”. So as my explanation made in (a), degree of procrastination should increase as these index increase.

The outcome variable ,Y, would be the same as the one in the model in (b).

The model is shown below, with β_0 being y-intercept(constant term), β_i being slope coefficients for each explanatory variable and ε being the model’s error term.

$$Y=\beta_0 +\beta_1 x_1 +\beta_2 x_2 +\beta_3 x_3 +\beta_4 x_4 +\beta_5 x_5 +\varepsilon$$

(d) For each of the two general models produced in (b) and (c), derive two interesting predictions (four predictions in total). Be sure the logical connection between your model and your predictions is explicitly stated and that any assumed facts concerning the world are made explicit.

For model in (b)， the first interesting prediction might be that the first variable "self-confidence" actually affects people's procrastination behavior in the opposite way of my expectation. as sometimes people with high confidence might start to work late because they believe they can finish the task in a relatively short period. The second interesting prediction could be that there is no statistically significant correlation between this variable of efficiency concern and the outcome variable. Since people value high efficiency may also have different strategies when doing tasks. Some may prefer to do it earlier so as to save time for other tasks, and some would choose to do it late to take advantage of the deadline pressure, so it is a working strategy choice rather than efficiency choice, and efficiency concern could not be a plausible explanation for procrastination.

For model in (c), one interesting prediction might be the negative correlation between level of difficulty and degree of procrastination, as people may tend to start the difficult task early rather than even lessening the possibility of solving problems out by putting off the work until last minute. Second, the variable of loads of other homework might interact with procrastination in an unexpected way rather than positive. It might turn out to be the case that moderate loads of assignments would make people less procrastinate as they want to save time for other assignments, but too little or too much assignments would both lead to procrastination, as the former means less pressure so one may start late, and the latter, on the other hand, brings too much pressure, or even results in mental health problems, rendering students impatient when facing so many tasks, and therefore postpone every assignment to a very late stage.

Selecting and fitting a model

1. For each part, indicate whether we would generally expect the performance of a flexible statistical learning method to be better or worse than an inflexible method. Justify your answer.

a. The sample size n is extremely large, and the number of predictors p is small.
A flexible method will perform better than an inflexible method, as an extremely

large sample size enables flexible models to avoid overfitting problem, fit the data closer and get an accurate estimate of f.

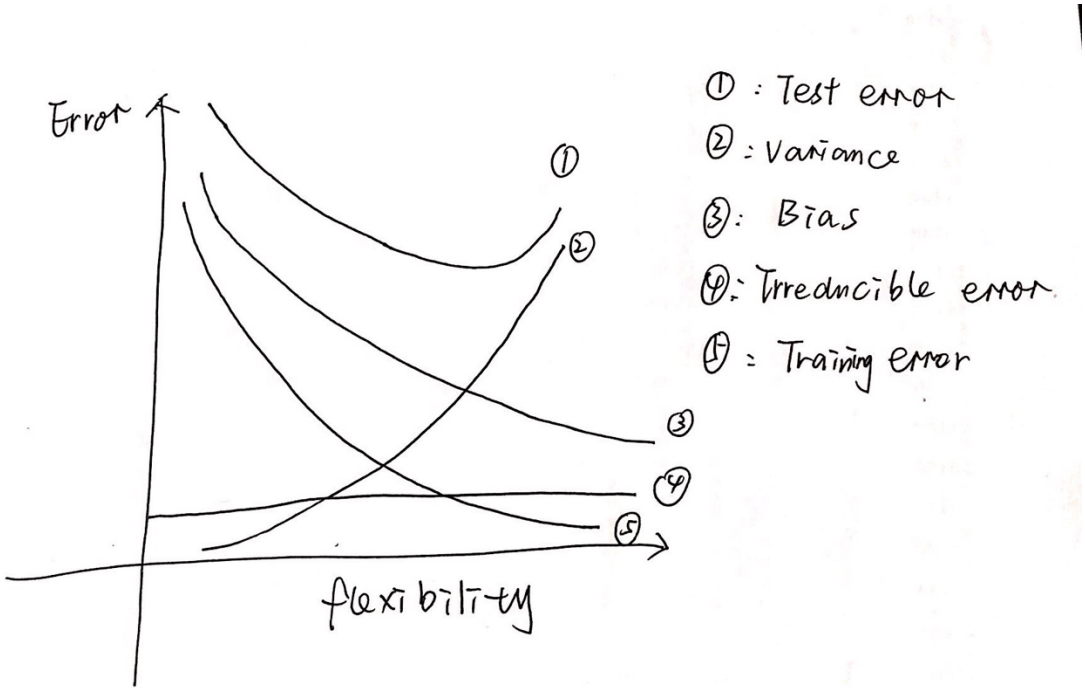
b. The number of predictors p is extremely large, and the number of observations n is small. A flexible method will perform worse than an inflexible method, since a small number of observations will expose us to overfitting problem, while an inflexible model can describe the true patterns better out of the noise

c. The relationship between the predictors and response is highly non-linear. A flexible method will perform better than an inflexible method better, because flexible methods have higher expressive power to generate a wider range of possible, especially non-linear shapes, while the inflexible method may not be able to capture the complex non-linear patterns and is more appropriate for linear relationships.

d. The variance of the error terms is extremely high. A flexible method will perform worse than an inflexible method, because it will fit to the error terms as well and increase variance.

2. Bias-variance: Think about the graph including bias, variance, training error, test error, and irreducible error curves, moving from less flexible statistical learning methods towards more flexible approaches.

a. Explain why each of the five curves has the shape it has.



Bias: Bias will decrease with the increase of flexibility. Bias rises from using simplified model, and using more flexible models allows us to better approximate complicated real-world "models" and situations, so bias would decrease.
Variance: Variance will increase with the increase of flexibility, because more flexible models follow the data more closely, so a small variation in any of the data points could significantly change the estimation.
Training error: Training error will decrease with the increase of flexibility because flexible models can fit the training data better.
Test error: Test error will change in a U-shape with the increase of flexibility. As we know, test error is the sum of squared bias, variance and irreducible error. When flexibility is low, variance is small and bias is large. The force of bias dominates and test error decreases with the increase of flexibility. When flexibility is high, however, variance is big and bias is small, the force of variance dominates and test error increases with the increase of flexibility.
Irreducible error: Irreducible error would not change with the increase of flexibility, because it is impossible for a model to perfectly describe the real data and irreducible error is the lowest achievable test error for all models.