Problem Set 1

MACS 30100 Winter 2020

Nak Won Rim

Building Models

Deviant Aggressive Behavior

1. What social policy might be appropriate to reduce deviant aggressive behavior if Theory I were correct? Theory II? Theory III? Theory IV? (Word Count: 574 words)

Theory I suggests that deviant aggressive behavior is the result of wrong social reinforcement learning. In other words, deviant aggressive behaviors are strongly related to positive gains but are not related to punishment strongly. Because of this, people learn to express deviant aggressive behaviors frequently. If we assume this theory gives the correct explanation about deviant aggressive behavior, an effective way to address it will be to reduce this social reinforcement learning. There could be two kinds of policies to reduce the reinforcement learning. One will be to strengthen the relationship between punishment and deviant aggressive behaviors. This could be achieved by imposing stronger punishment on deviant aggressive behavior, such as giving them more time in jail. Another way will be to weaken the relationship between positive outcomes and deviant aggressive behaviors. An example of this will be to educate the general population about alternative non-aggressive behaviors that are also strongly related to positive outcomes through the educational system. This will decrease the anticipated reward which the aggressive behavior gives relative to the desirable behaviors, ultimately weakening the relationship.

Theory II suggests that deviant aggressive behaviors are some sort of defense mechanisms that arise by displacing your anger toward yourself to figures outside of you. If this theory is correct, a direct policy addressing deviant aggressive behavior will be to provide more counseling support to people. Counseling will help people address their anger toward themselves and resolve their issues. One way to give people more counseling support will be to provide monetary support on counseling costs to people. Another way will be to launch a campaign that encourages people to see counselors, gradually building the social atmosphere that people could seek help when they are in need.

Theory III suggests that deviant aggressive behaviors are rational behaviors of people who do not benefit much from the social rules. Since following the social rules are not beneficial or even harmful for them, they do not conform to those rules, resulting in deviant aggressive behavior. If this theory is correct, policies aimed to increase the benefit of following social rules will be the most appropriate to decrease deviant aggressive behaviors. This will create more incentive for people to follow the social rules, ultimately decreasing the deviant aggressive behavior. A good example of this kind of policy will be to provide more social safety nets to people in the lower part of the socioeconomic gradient. Successful implementation of social safety net guarantees that the people in the worst condition gets support from the social structures to maintain a minimum life status. This will provide an incentive for the people to follow the social rules even in the worst situation.

Theory IV suggests that deviant aggressive behavior is a social role acquired from contacts with deviant subcultures. If an individual has frequent contact with deviant subcultures, such as organized crime gang, he will learn the deviant aggressive behavior as some kind of social role that is prevalent in society. If theory IV is correct in explaining the deviant aggressive behavior, an effective policy addressing the issue should focus on decreasing the social member's exposure to deviant subcultures. One way to achieve this will be to decrease the number of deviant subcultures people can be contacted with. This could be done by increasing the punishment for group activities expressing violent behaviors or by providing more options and incentives to people who wants to get out of such cultural groups.

Waiting until the last minute

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

One explanation for procrastination could be that people think they will get become more efficient in executing a task when the deadline for that task is near. On a personal example, I sometimes wander off to reading random articles or watching Youtube videos unrelated to the assignment while doing assignments when the deadline is far away, but this rarely happens when the deadline is near. Having more efficiency in doing a task will decrease the amount of time people have to spend to finish the task. Therefore, based on the assumption that people want to minimize the time allocated for tasks, people will wait until the last minute to gain maximum efficiency.

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.

A more generalized model of explanation provided in **a** could be that people predict their efficiency for executing tasks and start working at a time when they think efficiency could be maximized. As discussed before, time left before the deadline could be an important factor in predicting efficiency. However, there could be other factors involved, such as time of the day you are most active. For example, if you are an early bird who tend to function the best in the morning, your predicted efficiency will be higher in the morning. A linear model that takes into account two factors we discussed so far could be:

predicted efficiency $= \beta_0 + \beta_1 * ($*time to start working* - *deadline*$) + \beta_2 * |$*time to start working* - *optimal time of the day*$|$, $(\beta_2 < 0)$

and people start working at the time where predicted efficiency is maximized. This still explains the observation that people often procrastinate, but could be more generalized to other factors contributing to the predicted efficiency. I note that this model is based on the assumption that people predict their efficiency will be consistent throughout the time they work on the task. Also, the coefficient will vary by individuals, which will lead to predictions discussed in **d**.

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

An alternative model for explaining procrastination could be that people pick the action that gives you the most perceived utility at the moment, and the perceived utility of executing a task is affected by the time left before the deadline. If the perceived utility of executing a task is increased as we get closer to the deadline, at some point the perceived utility of executing the task will be greater than any other alternative activities. If that point is reached, people will start to do the task. Analogous to the simple model in b, a linear model for the perceived utility of doing a task could be:

perceived utility $= \beta_0 + \beta_1 * ($*current time* - *deadline*$) + \beta_2 * ($*perceived utility for finishing the task*$)$, $(\beta_1 < 0$, $\beta_2 > 0$ *and* $\beta_1 = 0$ *if the task has no deadline*)

This perceived utility could be compared to the perceived utility of alternative actions, and a person will choose to start this task only if the perceived utility is higher than any other task. I note that the coefficient will vary by individuals and by tasks, which will lead to predictions discussed in **d**.

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.

One prediction that could be derived from the model proposed in **b** is that if a person predicts the efficiency to decrease as it gets closer to the deadline, then that person will not procrastinate. In fact, the person would start as early as possible to maximize efficiency. This could potentially be due to increased panic as it gets closer to the deadline. Another way to describe this will be that for most people, the β_1 coefficient is negative, but for these individuals, the β_1 coefficient would be positive. Since β_1 is positive, the predicted efficiency will decrease as the deadline comes near for these people. Therefore, the time point where the predicted efficiency is maximized will be a time far before the deadline, leading to start early than to procrastinate. This could explain why some people start working on tasks earlier than other people.

Another prediction will be that if there are other factors strongly influencing the predicted efficacy, that person will be less influenced by the deadline. This could also explain why some people do not really procrastinate. Continuing with the factor of the optimal time of day discussed in **b**, if the absolute value of β_2 is much bigger than β_1 for a person, then the time which the predicted efficiency will be the greatest for that person will the time near her optimal time of day, not the time close to the deadline. Therefore, she will start working at the time near her optimal time of day instead of time close to the deadline. An example of such will be that if a person tends to work very well around 5 p.m. to 8 p.m but has terrible efficiency as the night gets deeper after that, then that person will try to start the task around 5 p.m. regardless of the deadline.

One prediction that could be derived from the model proposed in **c** is that if a person's life is mundane and does not have alternative activities with high utility, he will not procrastinate. This model proposes that a person will start working on the task if the perceived utility is relatively bigger than the perceived utility of alternatives, not when it passes a certain absolute threshold. Therefore, if he does not have some more attractive alternatives, he will not procrastinate and start working on the task faster.

Another prediction will be that if the utility for finishing the task is not high for the person, then he might not start the task even after the deadline. If the utility for doing a task is near 0, then the perceived utility for the task could be quite low even if the deadline approaches or even passes. If this is the case, then other attractive alternatives could still outweigh the perceived utility for doing the task and the person will not start the task. A good example of this behavior could be a person writing a recommendation letter for someone who he does not really care about.

I note that for both models, I assumed that people move rationally and could calculate abstract concepts such as predicted efficiency and perceived utility. Also, I assumed linearity and additivity for both models to simplify the explanation, but I think it is likely that other models could fit the real world better. For example, a polynomial term for (*time to start working* - *deadline*) could be better in the first model and having an interaction term between two variables in the second model could account for the real world better.

Selecting and fitting a model

1.

a. The sample size n is extremely large, and the number of predictors p is small.

A flexible model will perform better in this case. Since the number of parameter p is small, an inflexible model will likely be unable to explain the occurring pattern, leading to larger bias. A flexible model will capture the pattern better leading to better performance. In addition, if we assume that the data was sampled correctly, the large sample size will decrease the probability of overfitting. This could be also advantageous to using flexible models because the chance of overfitting is always a danger when applying flexible models.

b. The number of predictors p is extremely large, and the number of observations n is small.

An inflexible model will perform better in this case. Since the number of observations n is small, using a flexible model will be highly prone to overfitting. This will lead to the flexible model having poor performance in predicting the data outside the given observations. Also, since the number of parameters p is extremely large, even inflexible methods will show a relatively good fit for the data.

c. The relationship between the predictors and response is highly non-linear.

A flexible model will perform better in this case. An inflexible model will likely have a lower chance of grasping the non-linear relationships between predictors and response, especially if it is a model assuming linearity. Therefore, a flexible model will have less bias, leading to higher performance.

d. The variance of the error terms $\sigma^2 = Var(\epsilon)$ is extremely high.

An inflexible model will perform better in this case. A flexible model will try to find a pattern that explains the large variance the error terms bring in, leading to overfitting. An inflexible model will be less vulnerable to fitting the variance of the error terms, leading to better performance.

2. Bias-Variance

a. Explain why each of the five curves has the shape it has. (The figure is attached in the next page)

The bias (shown in the olive green curve) decreases as the flexibility of the model increases. The bias is essentially an error introduced by the difference between the model assumption and the real world. In other words, the bias will be greater if the model shows less fit to the data points. Flexible models allow better fitting to the data points, so the bias will decrease as the flexibility of the model increases.

The variance (shown in the blue curve) increases as the flexibility of the model increases. The variance is essentially how different each model will be when using different training data points. As mentioned before, flexible models tend to fit better to the training data points. Therefore, it will be more sensitive to fluctuation in the training data points, showing more variance. In other words, the noise in each data set will be better captured in the flexible models, leading to greater variance as the flexibility of the model increases.

The training error (shown in the black curve) decreases as the flexibility of the model increases. Flexible models tend to fit better to the given data points than inflexible models. This means that the prediction given by the flexible models will be closer to the actual values in the training set. Therefore, the training error will decrease as the model becomes more flexible.

The test error (shown in the green curve) decreases as the model becomes more flexible until a certain point, then increases as the model becomes more flexible. If the model is too inflexible, it will poorly predict the data points in both the training set and the test set. This is because the model assumption is too different from the actual phenomenon in the real world. If the model is too flexible, it will overfit to the training data points. It will fit the noises in the training data set, which will harm its generalizability to the data points outside the training set. Therefore, a model that is not too flexible nor too inflexible will have the lowest test error. This accounts for the shape of the curve the test error shows as the flexibility of the model increases.

Finally, the irreducible error (shown in the red line) stays constant as the flexibility of the model changes. Irreducible error is the error unrelated to the model and cannot be accounted for in any model. Therefore, an inflexible model will show the same degree of irreducible error with a flexible model. Because of this, the irreducible error stays constant as the flexibility of the model changes.

This article was written using R Markdown and was knitted to pdf using the tinytex package

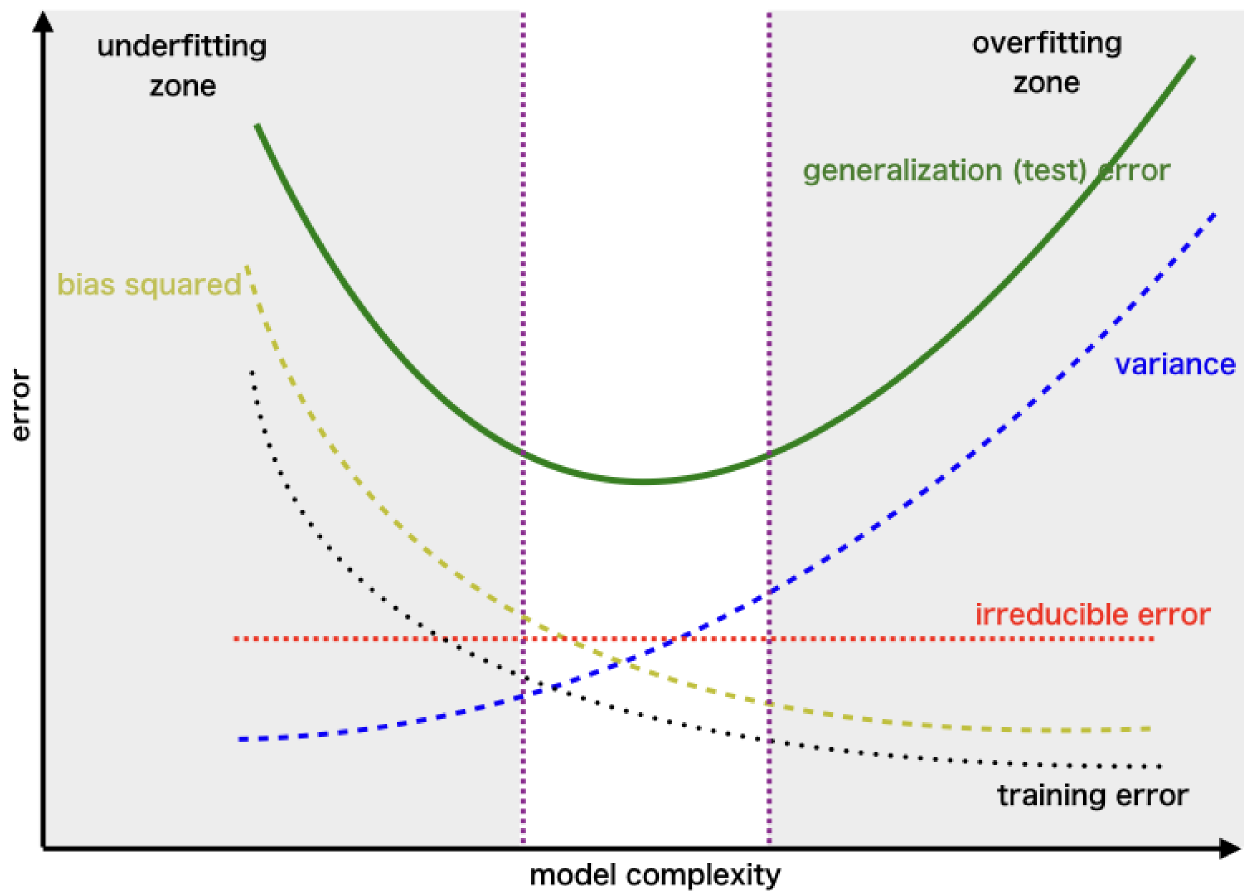

Figure 1: The five curves. Image from towards data science. Although the x-axis label says "model complexity", using model flexibility as the x-axis will give similar results.