MACS 30100 Homework 1

Benjamin Jun Hui Toh

January 14 2019

1 Building Models

1. Deviant aggressive behavior

    (a) If Theory I were the correct model for deviant aggressive behavior, then it follows that such behavior occurs because the reward system incentivizes them to do so. This implies that deviant aggressive behavior occurs due to lack of rewards for socially positive behavior, and inadequate punishment for negative behavior. A social policy that seeks to correct for this has to change the incentives that an individual faces when deciding whether to perform a particular action. Insofar as these actions are ingrained at an early age, in order for individuals to learn and inculcate these pro-social behavior, it is imperative to intervene at childhood. We can introduce education policies that have greater emphasis on civics and moral education, thus inculcating positive values in children. Schools can also award bursaries to children who reflect these pro-social behavior. To stigmatize anti-social behavior, there can be more stringent rules and tougher enforcement to ensure compliance. This can be applied to the judicial system as well, with harsher punishment to deter aggressive deviant behavior.

        If Theory II is correct, such deviant aggressive behavior is a form of lashing out by individuals against authority. This signifies a form of anti-establishment view of the world, and a social policy oriented around enforcement and punishment is unlikely to be effective. An effective social policy must, thus, center around ameliorating individuals' hostility and anger toward personal authority. Such an approach could incorporate counselling to resolve emotional and anger management issues. A community-based approach focused on providing emotional and financial support to these individuals would help them feel more integrated into the society and less angry in his personal life.

        Theory III implies that deviant aggressive behavior is a deliberate choice amongst individuals who feel oppressed by the society and hence feel that they have nothing to lose from such behavior. As such, any social policy that seeks to resolve this issue must tackle the prickly issue of discrimination and the sense of oppression that these individuals perceive. This could imply a need for affirmative action to correct for hitherto discrimination. For the younger generation, any social policy must seek to broaden the paths to success and work to ensure equality of opportunity for all. This could include a broader definition of meritocracy - when considering two possible candidates for jobs who come from different backgrounds, one might need to consider the relative difficulty both candidates had to experience when evaluating their performance. Social policy should also focus on community-based engagement and solutions, since community-based groups are more likely to be cognisant of problems faced by their community, and hence are better able to tailor support for these individuals.

        Theory IV suggests that deviant aggressive behavior is spread through culture, and that one learns such behavior from their exposure to deviant subculture. A social policy that seeks to correct for these might consider banning elements of such culture, or to step up educational campaigns about these subcultures and the anti-social behavior they encourage. Given culture is largely acquired through music, television, books and films, an appropriate social policy might also consider stepping up promotional efforts to encourage more pro-social culture as an alternative to these deviant subculture.

(b) Political rhetoric in America is a form of deviant aggressive behavior. Taking then-candidate Donald Trump with his divisive rhetoric on immigrants and the repeated assaults on the media as an example, it is a manifestation of Theory I in practice. Donald Trump's brash political comments received widespread media coverage, which served as free publicity for his campaign. In addition, his rhetoric endeared him to a hardcore base, whose loyalty to Trump was reflected in his comment that he could shoot somebody in 5th Avenue and not lose any voters. Indeed, there was little if no political blowback from his political rhetoric. This arguably helped to reinforce his deviant aggressive behavior, since he has learnt that his comments not only avoids any tangible punishment, but actually boosts his campaign's ratings. Indeed, the Trump campaign has showed the Theory I is a plausible explanation of such behavior.
Theory II is plausibly played out amongst Trump's voters. With the opioid epidemic plaguing many regions in the Rust Belt, there has been palpable frustration amongst these voters about their personal life, which manifests in the hostility toward authority. There is much anger about the inability of politicians to solve their problems, and thus turned to Trump, who had never run for office before and was perceived to be a candidate who "tells it like it is". This theory can plausibly explain Trump voters' steadfast support for candidate Trump regardless of the comments he made. This served to reinforce the fiery political rhetoric that the Trump campaign invoked in the lead up to the 2016 Presidential Elections.
A similar argument runs through Theory III as well. With Rust Belt voters facing poor economic prospects with the deindustrialization and closing of factories in the region, many Trump voters feel oppressed by the economic elite and the policies that they perceive to only benefit the rich and political elite, rather than themselves. They feel that they have been discriminated against through economic policies that politicians had implemented. As such, they feel rules drawn up by politicians merely entrench their own interests, and believe that there is little to lose by supporting a candidate that breaks all the norms. This can also be seen from the popularity of the slogan "Drain the swamp" during Trump rallies, implying a sense of frustration with current-era politics. The Trump election victory as seen from Theory III can, thus, be seen as a protest vote by voters who feel that they have nothing to lose by voting in a candidate who does not play by the rules.
Theory IV can possibly explain some part of how Trump voters started supporting the campaign. With the pernicious media influence of Fox News, and far-right sites such as Breitbart and Infowars, individuals who watch or browse these media outlets are fed a view that their country is under siege from immigrants and politicians, and the best way of getting out of this crisis is to vote for a candidate like Trump who is viewed as refreshingly non-politically correct. As such, Theory IV could explain the initiation of these voters into deviant aggressive behavior, participating in political rhetoric and chanting slogans such as "Lock her up!" during the campaign.

2. Waiting until the last minute

(a) The observation might be true due to frequent observations of individuals who to do so or incidents where such an event occurs. This can possibly be empirically verified through survey data. People often procrastinate due to time inconsistency between their current self and future self. An individual may care too much about the current self and inadequately about his or her future self. This might be a reason why people often do things at the last minute. Another reason why may be inertia bias, where individuals at rest stay at rest, until an external force such as imminent deadlines compel them to act. In addition, there is a lack of urgency in doing things early, whilst doing it last minute triggers a form of fight-or-flight response in people, which may enable them to overcome the inertia bias. There may also be distractions around the individual which result in him or her not attempting the work earlier, thus forcing them to do work at the last minute. There may also be self-control problems as an individual is unable to do the rational thing of finishing work before having fun, rather than procrastinating first and rushing work later.

(b) There may be a form of hyperbolic discounting for the individual between his current self and future self. An individual may discount future self's utility to a much greater degree than what future self would prefer. This would form the basis of a model of time-inconsistent preferences. A rational individual might discount future time periods using a particular discount rate, in a time-consistent manner. This would signify that time period 2 is discounted to time period 1 to the same extent as time period 3 discounted to time period 2. In a hyperbolic discounting model, events further away are discounted by a much larger rate than a time-consistent discounting model would predict. As such, rewards obtained at the end of work process provide relatively small utility compared to doing other fun things in the early time period. This explains why individuals will usually do fun things which deliver high utility, thus leaving work to the last minute.

(c) Inertia bias can be developed as model to explain this observation. Individuals may have a loss function for their utility, in which changing their current course of action may incur loss in utility. As such, individuals may prefer to remain in their present state, which is to not do their work. There is also loss from not being able to finish their work on time, and this loss increases closer to the deadline. At some point, with imminent deadlines, the loss incurred from not finishing their work on top dominates the loss from changing their current course of action. Therefore, individuals may then choose to change their course of action and do their work. This is the most acute immediately before the deadline, and hence people often do things at the last minute.

(d) For the model in (c), one interesting prediction is that individuals may make seemingly irrational choices to overcome the problem of hyperbolic discounting. For example, to overcome time-inconsistency problems of saving for retirement instead of spending today, individuals may choose to place a fixed portion of their monthly wages into a retirement account that cannot be withdrawn from until one hits retirement age. This is seemingly irrational as such an option is strictly dominated by a more flexible option that allows one to respond to financial shocks. Yet, individuals choose to do so because they are self-aware of their time-inconsistency problem, and thus choose such a plan to commit themselves to saving for retirement. Another interesting prediction of the model outside of procrastination would be to explain addiction. Smokers may be aware of the effects of smoking, but they hyperbolically discount their future self's health. As smoking provides them utility in the immediate time period, they continue to do so despite of awareness of future costs.

    For the model in (c), inertia bias can also provide another perspective to the addiction problem. Inertia bias would predict that addicts would remain on the current course of action(continuing to smoke) rather than changing course of action(quitting). This occurs until the point in which the loss of smoking dominates the loss from changing course of action. This may occur when smokers fall sick and the loss of utility of this sharply increases as a result. Another interesting prediction of the model is that many individuals who wish to change their jobs never end up doing so, as the loss of taking action is costly. Whilst these individuals may dislike their jobs and would gain higher utility of doing so, they would have to incur a loss of doing so. As such, inertia keeps individuals at a job that they dislike, in spite of awareness of potential (and possibly immediate) gains from switching jobs.

2 Selecting and fitting a model

1. (a) Since sample size n is large, flexible statistical learning methods allow for more parameters to be fit, so it performs better. Since n is much larger than p, this ensures asymptotic properties hold, and also reduces variance.
   (b) With small n and large p, flexible learning methods will do worse due to potential overfitting of the model.
   (c) Given the non-linear structure of data, a flexible statistical learning method would perform better since it makes fewer assumptions on data structure. A non-parametric model would perform better than a parametric (and worse if mis-specfied) model in this case.
   (d) If variance of error terms is large, then with a different training set, one might get an entirely different fit to the model. As such, flexible statistical methods do worse since they are more likely to overfit to the particular training data.
2. (a) I have attached the plot and the code verbatim below.

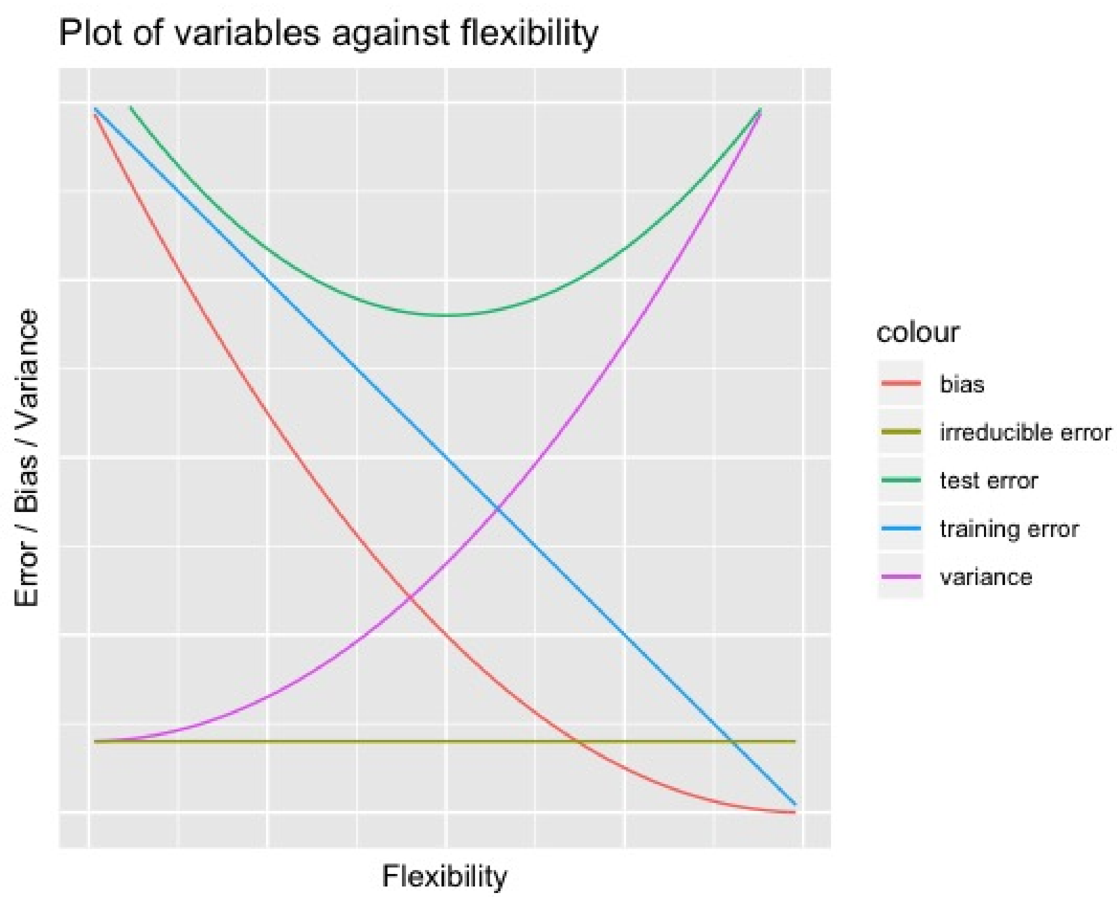

Figure 1: Plot of model flexibility against bias / error / variance

```
    library(tidyverse)

y1 = function(x){(x-10)^2} #bias
y2 = function(x){10 + x ^ 2} #variance
y3 = function(x){100 - 10 * x} #training error
y4 = function(x){70+ 1.5 * (x-5)^2} #test error
```

```
y5 = function(x){10} #irreducible error


set.seed(1492)

df <- data.frame(
  x = runif(100, min = 0, max = 10)
)
x <- df$x

p <- ggplot(data = df, mapping = aes(x = x))

p + stat_function(fun = y1, aes(color = "bias")) +
  stat_function(fun = y2, aes(color = "variance")) +
  stat_function(fun = y3, aes(color = "training error")) +
  stat_function(fun = y4, aes(color = "test error")) +
  stat_function(fun = y5, aes(color = "irreducible error")) +
  scale_y_continuous(limits = c(0,100)) +
  labs(x = "Flexibility", y = "Error / Bias / Variance",
  title = "Plot of variables against flexibility") +
  theme(
    axis.text.x = element_blank(),
    axis.text.y = element_blank(),
    axis.ticks = element_blank())
```

(b) Bias falls as the flexibility of the models increases because there are fewer assumptions made about the structure of the model, so the difference between actual and estimated parameter falls.

Irreducible error remains constant across the flexibility of the model since it is the error that cannot be avoided regardless of how well the parameter is estimated.

Test error shows a 'U' shaped parabola, with error initially falling as flexibility increases, before reaching a minimum point. After the minimum point, test error rises again as flexibility increases. This is because with a more complicated model, more of the variation in the test set can be explained with the training model. This occurs up to the point where the training model is overfit such that the model predicts training data very well, but does not generalize to the test set.

Training error falls as flexibility of model increases since including more and more variables can explain larger variation in the training set. Taken to its extreme, the model can perfectly fit training data, so training error falls to zero.

Variance increases when flexibility increases since the model would increasingly fit the specific training data set better, but the estimate may change by an increasingly large amount when a different training data set is used.

3.

```

#part a
set.seed(1492)

#part b
x1 = runif(200, min = -1, max = 1)
x2 = runif(200, min = -1, max = 1)

epsilon = rnorm(200, mean = 0, sd = sqrt(0.25))
```

```
#part c
y = function(x1,x2) {
  x1 + x1^2 + x2 + x2^2
}

#part d
logoddstoprob <- function(x){
  odds <- exp(x)
  prob <- odds / (1 + odds)
}

data <- tibble(x1, x2)
data <- mutate(data, logodds = y(x1, x2),
         probability = logoddstoprob(logodds),
         class = probability > 0.5
        )

sim_data <- tibble(x1, x2)

sim_data <- mutate(sim_data, logodds = y(x1, x2) + epsilon,
         Y = logodds > 0.5,
         probability = logoddstoprob(logodds),
         class = probability > 0.5
  )

#parts e and f
add_new_plot <- function(plot, df, ...){
  plot +
    geom_point(data = df, aes(x = x1, y = x2, color = class))
}

plot1 <- ggplot(data = data, aes(x1, x2, color = class)) +
  geom_density2d(data=data, aes(group = class))  #+ #geom_point()
plot1

plot2 <- plot1 %>% add_new_plot(sim_data)
plot2

#part g
plot2 + labs(title = "Bayesian Decision Boundary")
```

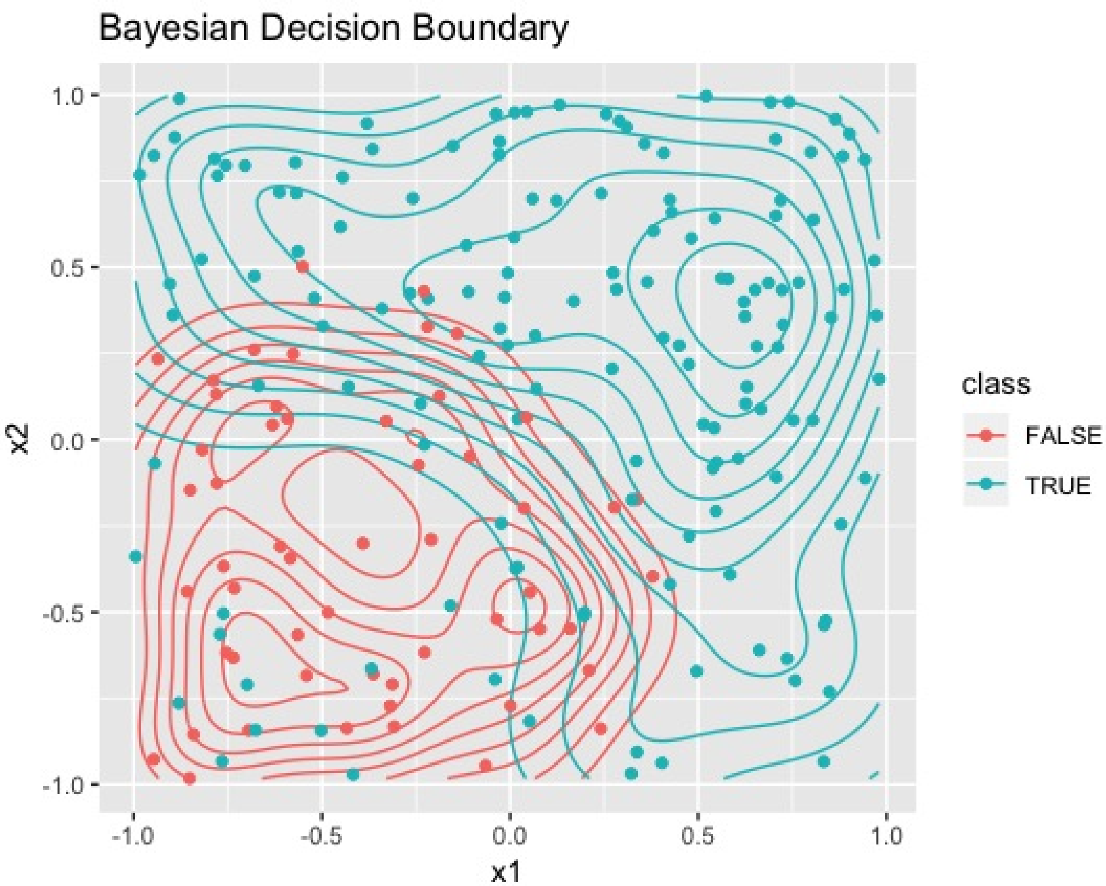

Figure 2: Plot of Bayesian Decision Boundary

p.s. sorry I know this wasn't the chart I had to plot but I was unable to plot the Bayesian Boundary Decision, so using geom_contour was the closest I could do.