

# Selecting and Fitting a Model

January 14, 2019

## 1 Building models

### 1.1 Deviant aggressive behavior

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

- Theory One: When people find practicing aggressive behavior potentially rewarding, they will naturally turn more inclined to obtain these rewards. Possible resources of such 'rewards' might stem from: 1) The feeling of being distinguished from the peers, and the reward of making life less 'mundane'. This might be due to the influence of entertainment movies or other misleading paper media. 2) The actual material and immaterial reward that one can acquire from such deeds, such as the power over others or that one can threaten others for money. **To mitigate such deviant aggressive behavior**, we can try policies to alter people's expectation of such rewards: for example, implementing stricter film rating systems especially online, to diminish exposure to violent movies. On the other hand, we can increase penalty to aggressive behavior, so that such rude manners are likely to entail more risk and less benefit.
- Theory Two: People might vent their anger, fury and hostility against the authority body through such aggressive behavior. This might be caused by the formidable gap there between these mad agents and the authorities over them. **To mitigate such deviant aggressive behavior**, we have to bridge the gap and reduce hierarchy within different levels of social institutions: for example, in companies, we will demand more democracy in all levels of management, require more staff engagement in decision-making processes, and establish more connection between the trade unions and the corporate; at the family level, we will encourage more communication between parents and children, require parents to accompany their children under certain ages for some amount of time; more publication of government work report, more involvement at online social platforms may also help to ease the tension between the general public and public officials.
- Theory Three: The most discriminated bodies are likely to take outrageous acts to attract more attention, bargain for more power and benefits, which is totally understandable in some sense. **To mitigate such deviant aggressive behavior**, we must make great efforts to secure social fairness and refrain systematic discrimination. Such policies could be concretized in a more economic way: for example, there is employment discrimination against women, as they might be pregnant someday and have to be absent from work for a while. To cope with such discrimination, we can allow companies to apply for extra tax allowance

when their female staff gets pregnant, so that for the companies there are certain profits to earn to employ female staff.

- Theory Four: there is some possibility that undesired social norms are formed, and people are stuck in that local equilibrium. **To mitigate deviant aggressive behavior**, such local equilibria have to be distorted and shifted to healthier steady states: for example, some senior school children may bully the younger ones to show off their seniority. Generally we can pose heavier penalty on such deeds, and raise awards for efforts containing such behavior. In that way, students can learn the risks in doing, and the benefits in not doing.

**1.1.2 During the past ten years, American society has been running a series of “experiments” with deviant aggressive behavior. Take any one of these experiments (e.g. #MeToo, mass shootings, political rhetoric) and discuss what we have learned about the four theories from this series of experiments.**

I will try to dissect **mass shootings** and see how the "experiment" could be explained by the four theories. According to TIME, mass shooting victims recently peaked in 2017, with 112 people killed and 531 wounded (Chris, 2017). In 2018, mass shootings are still persistent and inducing massive social anxiety. A good number of hypotheses have been proposed to explain why this is happening, such as "higher accessibility and ownership of guns", "mental illness" and "failure of background checks". Also, we can analyze the problem from a social science perspective:

- Shootings may be an irrational expression of anger by the chronically bullied bodies against other more powerful individuals. This explanation is a blend of Theory Two and Three. These violent criminals desperately need an outlet to shoot their fury, especially against authorities and those who stay aloof to their suffers. The act is akin to taking revenge. When they view the entire society as the foe, they trigger mass shootings. Psychological evidence has alluded that many such criminals have more or less experienced bullies during childhood, and that experience can easily translate into fatal consequences. (Burgess et al, 2006).
- There were reports that mass shootings are associated with individualistic culture. "Mass shooters in any nation tend to be loners with not much social support who strike out at their communities, schools and families", according to Prof. Peter Squires (Dorell, 2012). Part of the individualism is due to socioeconomic factors, such as economic and social strata segregation. This argument can partially be derived from Theory Four, which basically speaks about the influence of subcultures.
- When talking about Theory One, I argued that rewards associated with deviant aggressive behavior, such as mass shootings, can be in immaterial forms. One explanation can stand on this view, that some criminals intentionally desire for fame and notoriety. Anti-social activism is being praised and propagated across the Web. As Kim put it, "The prevalence of these crimes has risen in relation to the mass media coverage of them and the proliferation of social media sites that tend to glorify the shooters and downplay the victims." (2016) The desire for violence is another source that stimulates mass shootings.

As I argued, reducing gun violence is not solely about gun control, but demands for a thorough investigation of social problems. We should always bear in mind social science theories, and be able to deduce and identify potential perils out there, such as discrimination against ethnicity, violent subculture and effects of misleading mass media.

**Reference:**

- Chris Wilson (November 5 2017). 35 Years of Mass Shootings in the U.S. in One Chart. *TIME*. Retrieved from <http://time.com/4965022/deadliest-mass-shooting-us-history/>\*\*
- Wikipedia. Mass shootings in the United States. Retrieved from [https://en.wikipedia.org/wiki/Mass\\_shootings\\_in\\_the\\_United\\_States](https://en.wikipedia.org/wiki/Mass_shootings_in_the_United_States)\*\*
- Eric Madfis. (2017) In Search of Meaning: Are School Rampage Shootings Random and Senseless Violence?. *The Journal of Psychology* 151:1, pages 21-35.
- Oren Dorell (Dec. 11 2012). In Europe, fewer mass killings due to culture not guns. *USA TODAY*. Retrieved from <https://www.usatoday.com/story/news/world/2012/12/17/guns-mass-killings-worldwide/1776191/>

## 1.2 Waiting Until Last Minute

### 1.2.1 Ask yourself why the observation might be true and write down your explanations.

Procrastination is incredibly prevalent and is frequently observed in all types of circumstances. Take homework submission of MACSS 33000 as an example, it can be clearly seen that most students push their answers to the remote repository quite late. Someone might argue that such delayed submission can spare more time for students to make some last minute modification to their homework, so that they can hit satisfactory scores in the end. But in fact, people even retard the completion of simple tasks. For example, on weekends, people often wait until last minutes to pick up packages before the service is closed, while picking up packages earlier does not impose any more risk or effort. Therefore, observations of late acts are truly reflective of people's internal tendency of procrastination. I've come up with a simple idea to explain the phenomena, and I would like to name it as Cost Minimization Theory: for one thing, people are simply averse to fulfilling certain tasks, in other words, taking actions to fulfill these tasks will induce expectation of negative utility; for another, not taking the initiative at a time point will also induce expectation of negative utility, which is basically the risk of failing to fulfill the task on time. People compare these to cost expectation at each time point and make the decision to start when the expected cost of failure exceeds the expected cost of taking actions.

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

Here I will simply specify the endogenous variable, the exogenous variables, and the general function form how exogenous variables determine the outcome:

- Endogenous variable (outcome of interest): the decision of starting work at the moment. Denote it as  $Y$ , with  $Y = 1$  as deciding to start and  $Y = 0$  as deciding to wait.
- Exogenous variables (predictors): 1) the expected cost of completing the work, denoted as  $C_{do}$  which varies across individuals who have different skills and preference. 2) Given the cost of failure as prespecified, the expected cost of not starting the work at the moment, which is determined by the remaining time prior to the deadline. And I would love to denote it as  $C_{delay}(t)$ .  $C_{delay}$  is a monotonically decreasing function with regards to  $t$ , because the risk of failing the work increases as the deadline comes closer.
- The function form is quite intuitive, which simply compares the two costs and go for the minimal one.  $Y = I(C_{delay}(t) > C_{do})$ , where  $I(\cdot)$  is an indicator function.

This model setup should correspond to our findings in real life. It explains why people procrastinate, people only start work on stuff they feel adverse to when the expected cost of delaying surpasses the expected cost to do it; It also explains why people differ in procrastinating behavior, which is basically because each individual has a unique  $C_{do}$ .

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

To induce an alternative model, I would like to propose another mechanism that might cause procrastination. There are two assumptions: 1) People prefer to smoothe the efforts taken to complete certain tasks. Or more precisely, they often prefer to y schedule the task at the middle of their timeline, so that they can have free time at the beginning and the ending, and having their efforts smoothed. 2) People tend to overestimate the time left before the deadline. Then here is the model specification:

- Endogenous variable (outcome of interest): the decision of starting work at the moment. Denote it as  $Y$ , with  $Y = 1$  as deciding to start and  $Y = 0$  as deciding to wait.
- Exogenous variables (predictors): Let's take the estimated time needed to complete the task as given, and denote it as  $T$ . The remaining time prior to the deadline, denoted as  $t$ . There is another function of  $t$ , which determines the amount of time overestimated, denoted as the error term  $\epsilon(t)$  ( $\epsilon > 0$ ).
- The determining function form is also simple, people will start doing the task when their expected remaining time equals the expected time needed to complete the task. Precisely,  $Y = I(t + \epsilon(t) = T)$ , where  $I(\cdot)$  is an indicator function.

### 1.2.4 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 the model in (b), I would assume that people's expectation of two costs is stable and accurate. Under such conditions, people are pretty predictive of how much time and effort it would take to fulfill a certain task. Meanwhile, they are very good at estimating the probability of failing to complete the task at any time before the deadline. Then we can derive two very interesting results:

- Given a certain task, people's choice of starting time is stable. As assumed, people have consistent estimates of the two forms of costs, the solution of the determining function is stable then.
- People almost never fail to complete the task in the end. As assumed, people are good at estimating the costs, they will hence always find the right time to start work. If the cost of doing the task is always bigger, then they would have rejected the task in the first place.

For the model in (c), I would assume that people's extent of overestimation can be improved, which means holding  $t$  constant,  $\epsilon(t)$  can be reduced when the task repeats itself. Then we have the following two predictions:

- People sometimes fail to complete the task. When the task is new to them, their  $\epsilon$  turn to be excessive and they leave themselves no enough time to complete the task.

- People's choice of start time is different each time. As people can update their  $\epsilon(t)$  function each time they repeat the task, and the function can be learned from previous failures.

## 2 Selecting and Fitting a Model

**2.1 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.**

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

Under the circumstance, I would expect the flexible model to be more effective. Since we have a good number of observations, taking advantage of the sample size, we can fit the data better with flexible model settings. Inflexible models, like linear regressions, are likely to fail to match with the data especially when the relationship between features and responses are nonlinear.

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

If there are too few observations but a rich set of predictors, inflexible models are more likely to perform better. Since we do not have sufficient information about the data, it is tough to reach reliable predictions based on the data itself. But if we have any priori information about the mechanism between these predictors and the outcome, we can make full use of the predictors to derive good predictions.

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

If the true relationship is highly non-linear, flexible models would be the better choice. Inflexible methods usually assume for linear function forms, like LASSO and linear regressions. Methods based on these linear model specifications will lead to biased predictions, while more flexible models can better estimate the true relationship.

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

When there is hefty noise, I think inflexible models may perform better. Flexible models are inclined to overly fit the noise and fail to extract the function form. Inflexible models are better at fitting towards the function when we have the right set of predictors

## 2.2 Bias-Variance Decomposition

**2.2.1 a. Generate a graph of typical (squared) bias, variance, training error, test error, and Bayes (irreducible) error curves, on a single plot, as we go from less flexible statistical learning methods towards more flexible approaches. The  $x$ -axis should represent the amount of flexibility in the method, and the  $y$ -axis should represent the values for each curve. There should be five curves. Make sure to label each one.**

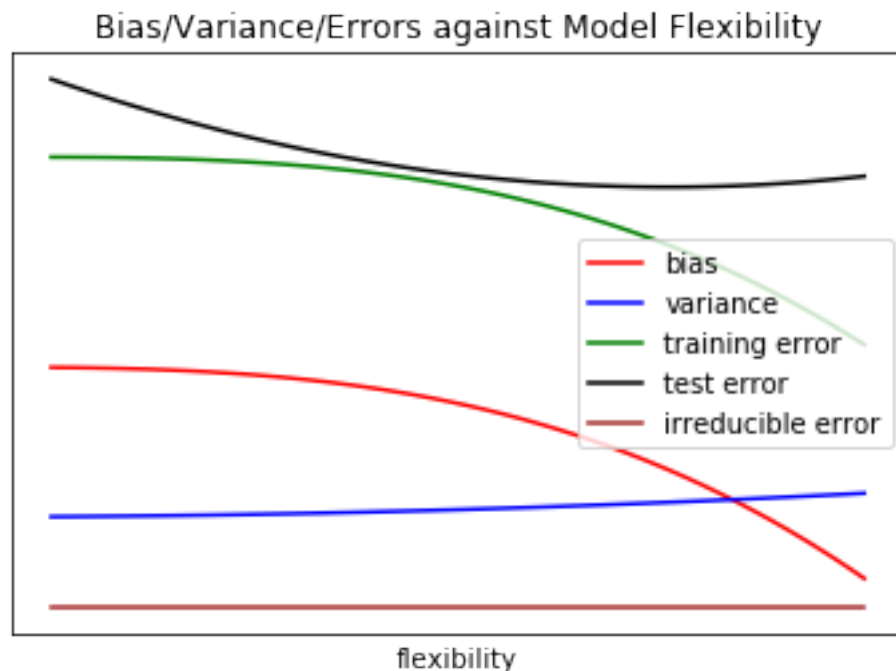
```
In [79]: # Import initial packages
import numpy as np
import matplotlib.pyplot as plt
```

```
In [120]: # Simulate data to approxiamte the curve
```

```
# x-axis
flexibility = np.arange(1, 9, 0.1)
# y-axis
# irreducible
bayes_err = [20] * len(flexibility)
# bias
variance = 0.1*flexibility**2 + 50
# variance
bias = -0.1*(flexibility)**3 + 100
# training error
train_err = bias + variance + bayes_err
# test error
test_err = (flexibility - 7)**2 + 160
```

```
In [121]: # Plot the graph
```

```
plt.figure()
plt.plot(flexibility, bias, color = 'red', label = 'bias')
plt.plot(flexibility, variance, color = 'blue', label = "variance")
plt.plot(flexibility, train_err, color = 'green', label = 'training error')
plt.plot(flexibility, test_err, color = 'black', label = 'test error')
plt.plot(flexibility, bayes_err, color = 'brown', label = 'irreducible error')
plt.xlabel("flexibility")
plt.xticks([]); plt.yticks([])
plt.legend()
plt.title("Bias/Variance/Errors against Model Flexibility")
plt.show()
```



### 2.2.2 b. Explain why each of the five curves has the shape displayed in part (a).

- The curve of the bias decreasing. With more flexibility in the model, data can be fitted better in the training set.
- The variance is increasing. With higher flexibility, local estimates depends more on the data instead of depending on the function. When dataset change, estimates are likely to change as well, which indicates that we have high variance in our estimate.
- The irreducible error should be a constant equal to the variance of the error term, which cannot be reduced altering model specification.
- The training error should be similar to the sum of bias(squared), variance and the irreducible error. Its shape should probably be downward, as flexible models are better at fitting in-sample data, which implies that variance increases slower than bias decreases.
- The test error is a convex curvature. When bias gets large, flexible models fit too much on sample noise and become less effective making out-of-sample predictions. Therefore, test error will raise on the right end of X-axis.

## 2.3 Bayes Classifier

```
In [84]: # Import initial packages
import numpy as np
import matplotlib.pyplot as plt
import math
```

```
In [85]: # set random seed
np.random.seed(2019)
# generate random data
df = np.random.uniform(-1, 1, (2, 200))
X1 = df[0,]; X2 = df[1,]
```

```
In [86]: # Calculate log-odds
Y = X1 + X1**2 + X2 + X2**2 + np.random.normal(0, 0.5, 200)
```

```
In [87]: # Use the log-odds to calculate probabilities
prob = 1/(1 + np.exp(-Y))
```

```
In [88]: idx = (prob > 0.5)
pos_x = X1[idx]; pos_y = X2[idx]
neg_x = X1[idx==False]; neg_y = X2[idx==False]
```

```
In [89]: # Calculate the boundaries (The boundary lies at where prob=0.5, which is identical to Y=0)
bound1 = []; bound2 = []; x = []
for x1 in np.arange(-1, 1, 0.001):
    a = x1 + x1**2
```

```

if a>0.25: break
y1 = (-1 + math.sqrt(1-4*a))/2
y2 = (-1 - math.sqrt(1-4*a))/2
bound1.append(y1); bound2.append(y2); x.append(x1)

```

```

In [92]: plt.figure("classifier")
plt.scatter(pos_x, pos_y, color = 'yellow', alpha = 0.5, label = 'success')
plt.scatter(neg_x, neg_y, color = 'blue', alpha = 0.5, label = 'failure')
plt.plot(x, bound1, color = 'red')
plt.plot(x, bound2, color = 'red')
plt.ylim(-1,1)
plt.xlabel("X1"); plt.ylabel("X2")
plt.legend()
plt.title("Bayes Classifier")
plt.show()

```

