

Name: Chun-Chi Huang
Course: Data Modeling
Professor: Michael McKee
Date: 02/01/2018

Chapter 3

4. In Chapter 3 of Predictive Analytics there are 4 quotes or sections. I would like you to pick one, and explain in your own words what that means:

"Correlation Does Not Imply Causation"

After looking at many examples in this chapter, I consider that PA is a powerful tool to find the correlations behind things. Sometimes, we can discover the pattern between two facts. The higher possibility of the first thing occurring, the more likely the second thing will happen. However, this does not mean that these two things have a direct relationship, and it may be just a coincidence. Take the example from this chapter: "Increased ice cream sales correspond with increased shark attacks". This first explanation in this chapter is funny, because it sounds like sharks prefer to attack people who eat ice cream because ice cream make people taste delicious for sharks. I do not agree with this explanation because it does not make sense. However, the second explanation sounds more reasonable: Due to warm weather, more people eat ice cream and go to the beach, and then the possibility of being attacked by sharks is in this warm weather. The chapter also mentions the example that "smokers get less carpal tunnel syndrome". This explanation seems reasonable because smokers need to smoke several times a day due to addiction, so they have to regularly take a break from using computers. My own example of PA is that athletes seldom suffer carpal tunnel syndrome. This may be because athletes regularly exercise which helps them avoid carpal tunnel syndrome, or because athletes seldom have a chance to use computers for a long time due to heavy training.

This topic also reminds me of my previous project in my bachelor degree. My goal was to use near infrared ray to predict the sweetness of mangoes. First, we scanned mangoes to get the absorption spectrum of the near infrared ray, and then physically measure the sweetness value of the flesh of the mangoes by using a sugar meter. Finally, we found a correlation between the absorption spectrum and sweetness value. When we only used single absorption of a specific wavelength to do regression analysis, there was a 60% accuracy. Moreover, when we used three wavelengths, we achieved almost 90% accuracy. Using this model, we could precisely predict the sweetness of mangoes without physically cutting and damaging them. During the experiment, I also found that redder mangoes were generally sweeter than normal ones. Maybe this observation could be considered another PA.