MATH 342W / 650.4 / RM742 Spring 2024 HW #1

Osman Khan

Monday 12$^{\text{th}}$ February, 2024

Problem 1

These are questions about Silver's book, the introduction and chapter 1.

(a) [easy] What is the difference between *predict* and *forecast*? Are these two terms used interchangably today?

To predict is to make a statement about the future (possibly from a spiritual source or one's mood), while forecast is to guess based on analysis. Today, both terms are used interchangeably.

(b) [easy] What is John P. Ioannidis's findings and what are its implications?

Ioannidis' findings is that there is not much carryover from the success of a medical trial in laboratory conditions to the reality outside. The implications are that we have to take claims about medical trials with a pinch of salt.

(c) [easy] What are the human being's most powerful defense (according to Silver)? Answer using the language from class.

Seeing patterns, making generalizations based on them, and then to respond to them. This means humans can use settings to forecast phenomena.

(d) [easy] Information is increasing at a rapid pace, but what is not increasing?

Objective Truth.

(e) [difficult] Silver admits that we will always be subjectively biased when making predictions. However, he believes there is an objective truth. In class, how did we describe the objective truth? Answer using notation from class i.e. $t, f, g, h^*, \delta, \epsilon, t, z_1, \ldots, z_t, \delta, \mathbb{D}, \mathcal{H}, \mathcal{A}, \mathcal{X}, \mathcal{Y}, X, y, n, p, x_{\cdot 1}, \ldots, x_{\cdot p}, x_{1\cdot}, \ldots, x_{n\cdot}$, etc.

$\mathcal{Y} = t(z_1, z_2, \cdots, z_n)$, a phenomena we wish to predict.

(f) [easy] In a nutshell, what is Karl Popper's (a famous philosopher of science) definition of *science*?

Testable hypotheses that, through experiments, are either i) proven wrong or ii) have not been proven wrong yet.

(g) [harder] Why did the ratings agencies say the probability of a CDO defaulting was 0.12% instead of the 28% that actually occurred? Answer using concepts from class.

The model used was based on information that was not completely relevant to CDO's, .i.e flawed and not enough historical data: $\mathbb{D}$. In addition, the model assumed that the mortgages were independent, when they actually were dependent, so $\mathcal{H}$ was flawed.

(h) [easy] What is the difference between *risk* and *uncertainty* according to Silver's definitions?

Risk is an easily quantifiable undesirable outcome while uncertainty is a not easily quantifiable undesirable outcome.

(i) [difficult] How does Silver define *out of sample*? Answer using notation from class i.e. $t, f, g, h^*, \delta, \epsilon, z_1, \ldots, z_t, \delta, \mathbb{D}, \mathcal{H}, \mathcal{A}, \mathcal{X}, \mathcal{Y}, X, y, n, p, x_{\cdot 1}, \ldots, x_{\cdot p}, x_{1\cdot}, \ldots, x_{n\cdot}$, etc. WARNING: Silver defines *out of sample* completely differently than the literature, than practitioners in industry and how we will define it in class in a month or so. We will explore what he is talking about in class in the future and we will term this concept differently, using the more widely accepted terminology. So please forget the phrase *out of sample* for now as we will introduce it later in class as something else. There will be other such terms in his book and I will provide this disclaimer at these appropriate times.

For Silver, out of sample meant using irrelavant historical data. For us, this means high ignorance i.e. δ.

(j) [harder] Look up *bias* and *variance* online or in a statistics textbook. Connect these concepts to Silver's terms *accuracy* and *precision*. This is another example of Silver using non-standard terminology.

Bias is accuracy and variance is precision.

Problem 2

Below are some questions about the theory of modeling.

(a) [easy] Redraw the illustration of Earth and the table-top globe except do not use the Earth and a table-top globe as examples (use another example). The quadrants are connected with arrows. Label these arrows appropriately.

(b) [easy] Pursuant to the fix in the previous question, how do we define *data* for the purposes of this class?

For the purposes of this class, data is natural result of phenomenon being measured.

(c) [easy] Pursuant to the fix in the previous question, how do we define *predictions* for the purposes of this class?

Predictions are the ability of a model to tell us what will happen in a certain phenomenon in a certain setting.

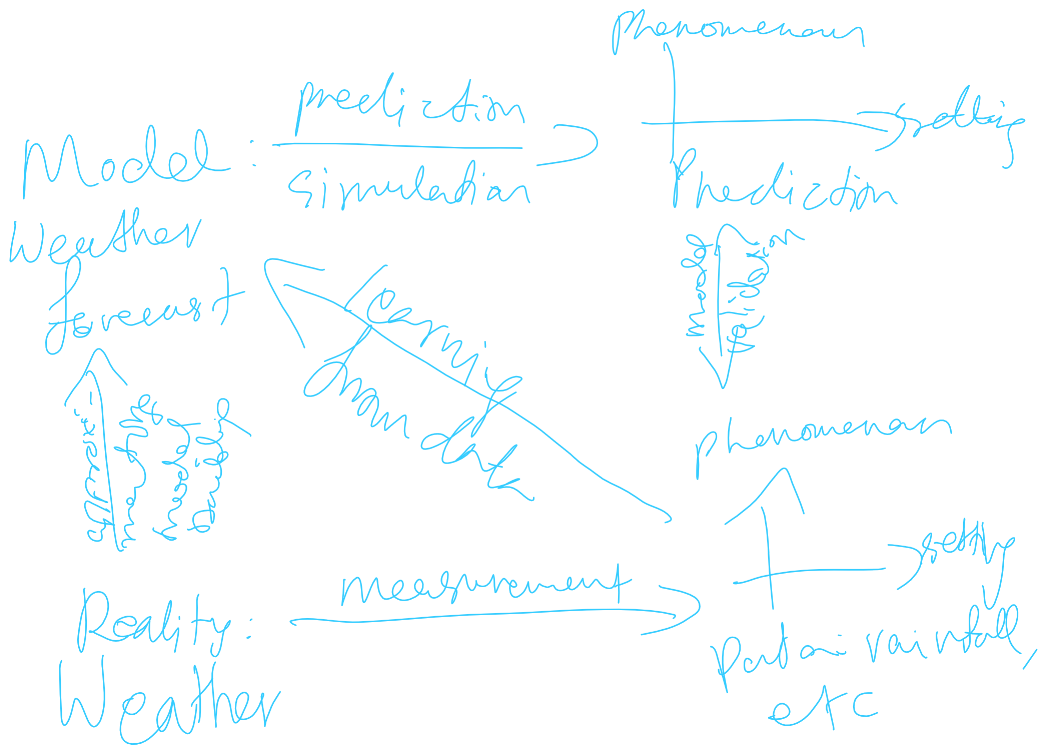

Figure 1: Model Valuechain

(d) [easy] Why are “all models wrong”? We are quoting the famous statisticians George Box and Norman Draper here.

All models are wrong as they are all approximations.

(e) [harder] Why are “[some models] useful”? We are quoting the famous statisticians George Box and Norman Draper here.

Some models are useful as despite their approximate nature, some models are close enough to the truth. We are able to work with what the model gives us.

(f) [harder] What is the difference between a "good model" and a "bad model"?

A good model is useful and could give us insights, about something simple. A bad model makes connections when there are none and/or is highly complicated and/or attempts to predict something very complex.

Problem 3

We are now going to investigate the famous English aphorism “an apple a day keeps the doctor away” as a model. We will use this as springboard to ask more questions about the framework of modeling we introduced in this class.

(a) [easy] Is this a mathematical model? Yes / no and why.

Yes, this is a mathematical model as it says that eating an apple everyday would mean never having to see the doctor.

(b) [easy] What is(are) the input(s) in this model?

An apple a day.

(c) [easy] What is(are) the output(s) in this model?

Good health.

(d) [harder] How good / bad do you think this model is and why?

This is a bad model. A fruit, like an apple, is just one small component of a healthy diet. There are numerous people who eat apples every day and still get sick and need to go to the doctor.

(e) [easy] Devise a metric for gauging the main input. Call this x_1 going forward.

Average number of apples eaten per day, over a year.

(f) [easy] Devise a metric for gauging the main output. Call this y going forward.

Number of times a person was sick enough to have needed to go to the doctor in a year.

(g) [easy] What is $\mathcal{Y}$ mathematically?

The number of times a person went to the doctor over a year.

(h) [easy] Briefly describe $z_1, \ldots, z_t$ in English where $y = t(z_1, \ldots, z_t)$ in this *phenomenon* (not *model*).

$z_1, \ldots, z_t$'s are the factors that determine a person's health. They could be genetic, dietary, environmental, etc.

(i) [easy] From this point on, you only observe x_1. What is the value of p?

1

(j) [harder] What is $\mathcal{X}$ mathematically? If your information contained in x_1 is non-numeric, you must coerce it to be numeric at this point.

$\mathcal{X}$ is the number of apples that subjects had on average on the course of a year.

(k) [easy] How did we term the functional relationship between y and x_1? Is it approximate or equals?

It is equal when we also include δ, otherwise it is approximate.

(l) [easy] Briefly describe *supervised learning.*

Using Training Data, Candidate Functions, and an Algorithem, supervised learning attempts to approximate f.

(m) [easy] Why is *superivised learning* an *empirical solution* and not an *analytic solution*?

An analytic solution can be arrived at through mathematics, even and especially in the absence of data.

(n) [harder] From this point on, assume we are involved in supervised learning to achieve the goal you stated in the previous question. Briefly describe what $\mathbb{D}$ would look like here.

$\mathbb{D} =< X, \vec{y} >$

X would be the average number of apples eaten in a year, while $\vec{y}$ would be the number of sbujects whose data regarding apple eating we were able to get.

(o) [harder] Briefly describe the role of $\mathcal{H}$ and $\mathcal{A}$ here.

$\mathcal{H}$ would be the list of candidate functions that can be used to approximate f. $\mathcal{A}$ is the algorithm that takes in the candidate function & historical data as inputs and gives us g, which is an approximation of f.

(p) [easy] If $g = \mathcal{A}(\mathbb{D}, \mathcal{H})$, what should the domain and range of g be?

The domain should be number of apples eaten and trips to the doctor over a year & the range should health (either trips to the doctor, or no trips).

(q) [easy] Is $g \in \mathcal{H}$? Why or why not?

Yes, as we defined $\mathcal{H}$ to be a set of candidate functions and g could very well belong to it.

(r) [easy] Given a never-before-seen value of x_1 which we denote x^*, what formula would we use to predict the corresponding value of the output? Denote this prediction $\hat{y}^*$.

$\hat{y}^* \approx g(x^*)$

(s) [harder] f is the function that is the best possible fit of the phenomenon given the covariates. We will unfortunately not be able to define "best" until later in the course. But you can think of it as a device that extracts all possible information from the covariates and whatever is left over δ is due exclusively to information you do not have. Is it reasonable to assume $f \in \mathcal{H}$? Why or why not?

No, as f is arbitrarily complex while those in $\mathcal{H}$ are simple.

(t) [easy] In the general modeling setup, if $f \notin \mathcal{H}$, what are the three sources of error? Copy the equation from the class notes. Denote the names of each error and provide a sentence explanation of each. Denote also e and $\mathcal{E}$ using underbraces / overbraces.

The three sources of error are: ignorance (δ), model misspecification, & estimation error. The sum of them is the residual error, denoted e.

$$y = g(\vec{x}) + \underbrace{h^*(\vec{x}) - g(\vec{x}) + \overbrace{f(\vec{x}) - h^*(\vec{x}) + \underbrace{t(\vec{z}) - f(\vec{x})}_{\delta\text{: } ignorance\ error}}^{\mathcal{E}\text{: } model\ misspecification\ error\ +\ ignorance\ error}}_{e\text{: } estimation\ error\ \mathcal{E}}$$

(u) [easy] In the general modeling setup, for each of the three source of error, explain what you would do to reduce the source of error as best as you can.

The ignorance error can be reduced by having relevant proxies, model misspecification error can be reduced by having more complicated models, while estimation error can be reduced by increasing the sample size.

(v) [harder] In the general modeling setup, make up an f, an h^* and a g and plot them on a graph of y vs x (assume $p = 1$). Indicate the sources of error on this plot (see last question). Which source of error is missing from the picture? Why?

Let $f = tan(x)$ (green), g be linear models (purple), $h^*(x) = 8(x - \frac{1}{2})$ (blue).

(w) [easy] What is a null model g_0? What data does it make use of? What data does it not make use of?

The null mode is what we have before incorporating any real data into our model. It uses $\vec{y}$ & does not use X.

(x) [easy] What is a parameter in $\mathcal{H}$?

λ

(y) [easy] Regardless of your answer to what $\mathcal{Y}$ was above in (g), we now coerce $\mathcal{Y} = \{0, 1\}$. What would the null model g_0 be and why?

The null model g_0 would be Mode$[\vec{y}]$

(z) [easy] Regardless of your answer to what $\mathcal{Y}$ was above in (g), we now coerce $\mathcal{Y} = \{0, 1\}$. If we use a threshold model, what would $\mathcal{H}$ be? What would the parameter(s) be?

$\mathcal{H} = \{\mathbb{1}_{x \geq \theta} : \theta \in \mathcal{X}\}$ The parameter is θ

(aa) [easy] Give an explicit example of g under the threshold model.

As from class: $\mathbb{1}_a = \begin{cases} 1 & \textit{if a is true,} \\ 0 & \textit{if a is false.} \end{cases}$

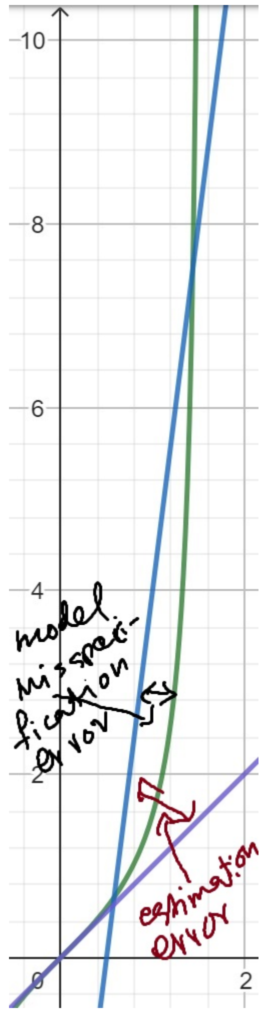

Figure 2: Ignorance error is missing from the picture as we do not know $t(\vec{z})$

Problem 4

As alluded to in class, modeling is synonymous with the entire enterprise of science.

In 1964, Richard Feynman, a famous physicist and public intellectual with an inimitably captivating presentation style, gave a series of seven lectures in 1964 at Cornell University on the "character of physical law". Here is a 10min excerpt of one of these lectures about the scientific method. Feel free to watch the entire clip, but for the purposes of this class, we are only interested in the following segments: 0:00-1:00 and 3:48-6:45.

(a) [harder] According to Feynman, how does the scientific method differ from learning from data with regards to building models for reality? (0:08)

For the scientific method, we first guess, compute consequences, conduct experiments, and then compare those computation results to nature/experiment/experience.

(b) [harder] He uses the phrase "compute consequences". What word did we use in class for "compute consequences"? This word also appears in your diagram in 2a. (0:14)

Approximation/Model Building

(c) [harder] When he says compare consequences to "experiment", what word did we use in class for "experiment"? This word also appears in your diagram in 2a. (0:29)

Prediction/Simulation

(d) [harder] When he says "compare consequences to experiment", which part of the diagram in 2a is that comparison?

Model Validation from Quadrant I to Quadrant IV.

(e) [difficult] When he says "if it disagrees with experiment, it's wrong" (0:44), would a data scientist agree/disagree? What would the data scientist further comment?

A data scientist would disagree. It is not necessarily wrong. If the inputs were relevant enough, we still could have learned something from model validation, and we would be able to improve out models.

(f) [difficult] [You can skip his UFO discussion as it belongs in a class on statistical inference on the topic of H_0 vs H_a which is *not* in the curriculum of this class.] He then goes on to say "We can disprove any definite theory. We never prove [a theory] right... We can only be sure we're wrong" (3:48 - 5:08). What does this mean about models in the context of our class?

The models can be useful for us, or it can be useless, but the model we develop for this class cannot tell us what definite truth is.

(g) [difficult] Further he says, "you cannot prove a *vague* theory wrong" (5:10 - 5:48). What does this mean in the context of mathematical models and metrics?

For mathematical models, you need specific metrics for inputs and outputs.

(h) [difficult] He then he continues with an example from psychology. Remeber in the 1960's psychoanalysis was very popular. What is his remedy for being able to prove the vague psychology theory right (5:49 - 6:29)?

Putting a metric on inputs and outputs will enable us to prove the vague psychology theory right.

(i) [difficult] He then says "then you can't claim to know anything about it" (6:40). Why can't you know anything about it?

As we do not know how much love is necessary, and we do not know how to measure love, therefore we cannot say what its impact would be on the upbringing of a child.

Just to demonstrate that this modeling enterprise is all over science (not just Physics), I present to you the controversial theoretical political scientist John

Mearsheimer. He's all over youtube and there's nothing special about this video that I will link here about Can China Rise Peacefully? Feel free to watch the entire clip, but for the purposes of this class, we are only interested in the following segments referenced in the questions which has nothing to do with China, only his theory of "power politics".

(j) [difficult] Is Mearsheimer's model of great power politics / international relations (i.e., modern history) 9:35-17:22 simple or complicated? Explain.

The model is simple, as there are five assumptions about countries, being: anarchic, military, intentions, survival, and rationality. But, countries are not real.

(k) [difficult] Summarize his ideas about limitations of his theory from 39:18-40:00 using vocabulary from this class.

The limitations on his theory is that the model at best is right 75% of the time. This would mean it could have low accuracy.