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

Professor Adam Kapelner

Due 11:59PM February 11, 2024 by email

(this document last updated 9:23am on Tuesday 6th February, 2024)

Instructions and Philosophy

The path to success in this class is to do many problems. Unlike other courses, exclusively doing reading(s) will not help. Coming to lecture is akin to watching workout videos; thinking about and solving problems on your own is the actual "working out." Feel free to "work out" with others; **I want you to work on this in groups.**

Reading is still *required*. For this homework set, read the first chapter of "Learning from Data" and the introduction and Chapter 1 of Silver's book. Of course, you should be googling and reading about all the concepts introduced in class online. This is your responsibility to supplement in-class with your own readings.

The problems below are color coded: green problems are considered *easy* and marked "[easy]"; yellow problems are considered *intermediate* and marked "[harder]", red problems are considered *difficult* and marked "[difficult]" and purple problems are extra credit. The *easy* problems are intended to be "giveaways" if you went to class. Do as much as you can of the others; I expect you to at least attempt the *difficult* problems.

This homework is worth 100 points but the point distribution will not be determined until after the due date. See syllabus for the policy on late homework.

Up to 7 points are given as a bonus if the homework is typed using LaTeX. Links to instaling LaTeX and program for compiling LaTeX is found on the syllabus. You are encouraged to use `overleaf.com`. If you are handing in homework this way, read the comments in the code; there are two lines to comment out and you should replace my name with yours and write your section. The easiest way to use overleaf is to copy the raw text from hwxx.tex and preamble.tex into two new overleaf tex files with the same name. If you are asked to make drawings, you can take a picture of your handwritten drawing and insert them as figures or leave space using the "\vspace" command and draw them in after printing or attach them stapled.

The document is available with spaces for you to write your answers. If not using LaTeX, print this document and write in your answers. I do not accept homeworks which are *not* on this printout. Keep this first page printed for your records.

NAME: Joshua Gisner

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? These 2 terms are "largely used interchangeably today."

A forecast was when one plans something under conditions of uncertainty.

A prediction is something you read out of a fortune cookie.

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

Most successful predictions that were carried out in medical labs were likely to fail when carried out in the real world.

It implies that it's hard to trust any hypothesis that aren't tested in the real world.

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

Our most powerful defense is identifying phenomena, and figuring out ways to measure and predict the outcome using settings from the real world.

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

Our ability to understand and process the information

(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.

The objective truth will be the phenomenon it's self that when we plug in our 3 z's, we get our exact output (which is impossible, that's why we approximate)

$y = t(z_1, z_2 \ldots z_t)$

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

That something is not scientific unless it can be tested in the real world via prediction

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

their rates were not derived from historical data, but just a bad statistical model.

Ther settings they used to measure the phenomenon were bad and ambiguous

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

Risk is when you make a calculated move based on the odds that you are given. while uncertainty is risk that is hard to measure. It's when your odds can be off by a factor of 100 or 1000 but there's no real way to know

(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.

It's kind of like cherry picking data, not looking at all the specifics to your situation and just looking at the big picture (or vice versa)

It's when your model, y, uses either too specific metrics or too broad of metrics to scope ~~the~~ phenominom

(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 and variance are prediction errors.

Bias is the difference bet. the avg prediction of our model, and the correct value we are trying to predict

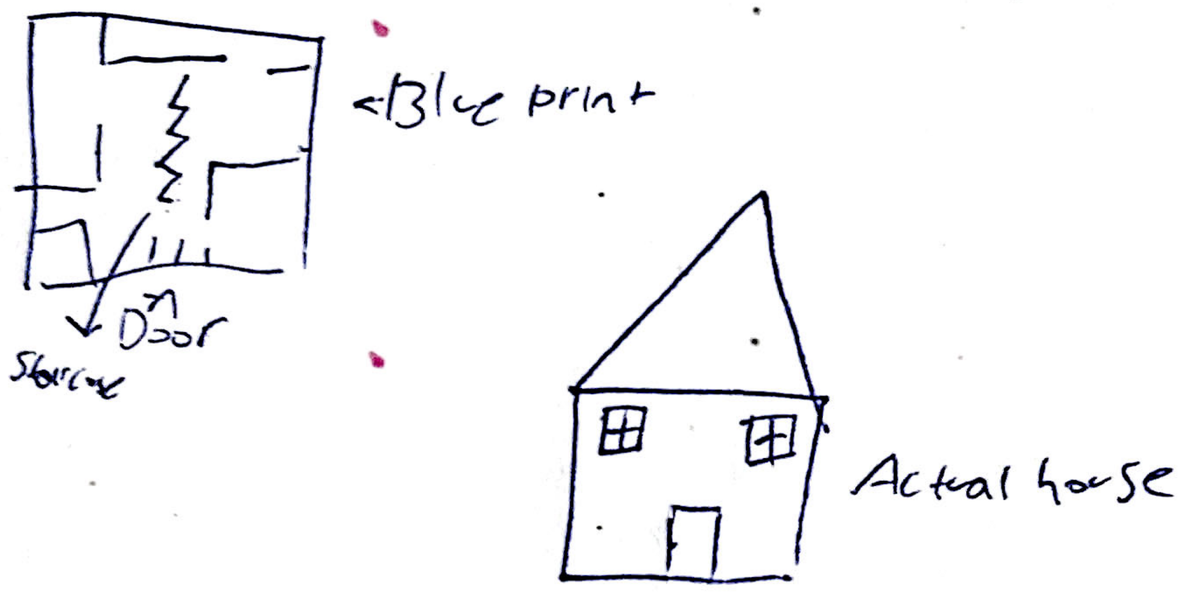
when a model has high bias it doesn't pay so much attention to training data and leads to a lot of errors. It is not accurate or precise

Variance does pay a lot of attention to the training data and doesn't generalize on data it hasn't seen before, so it performs very percisely on the training data, but it isn't nessecarily accurate bc it doesn't make up for the data it misses

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? As Settings and metrics

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

(d) [easy] Why are "all models wrong"? We are quoting the famous statisticians George Box and Norman Draper here. They are simplifications of the complex reality there are too many tiny variables that make it impossible to exactly model something. If it was exact, it wouldn't be a model, it would be the real thing

(e) [harder] Why are "[some models] useful"? We are quoting the famous statisticians George Box and Norman Draper here.
Bc although they are not perfect, we are able to gain some useful information from their outputs

(f) [harder] What is the difference between a "good model" and a "bad model"?
A good model provides accurate and precise representations of the real world. It's as simple as possible, and it's also able to perform off the data it hasn't seen yet

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.
No bc it's not modelled by numbers

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

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

The doctor who is "away"

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

The actual input is not good, just eating an apple daily will not keep you super healthy, However, if you view it as obtaining a healthy lifestyle everyday it's a pretty good way to remain healthy and in good shape and keeping yourself out of the doctors office

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

x_1 = amount of apples eaten daily

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

y = times per year seeing the doctor

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

y is a binary space saying he either did see the doctor or didn't.

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

the Z's are our inputs that would give us an exact way to measure a phenomenon

(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.

It's the space containing all of the approximating values we plug in for our Z's

In this case it will be a binary where 0 represents him not eating an apple and 1 will represent he did

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

we termed it as an approximate relationship, Z's would give us an equal relationship

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

a type of machine learning where an algo is trained using a dataset that includes inputdata and it's corresponding output. It's goal is for the algo to learn how to map the inputs to outputs and allow it to make accurate predictions for new, unseen data

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

It learns from data examples and doesn't derive solutions from theoretical or mathematical ideas.

(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. $\mathbb{D}$ would contain data inputs that have already been seen by the machine

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

$\mathcal{H}$ is being used bc we need to pick the best fc from the candidate space that contains the most appropriate data sets for our respective phenomenon

$\mathcal{A}$ is still our algo taking in $\mathbb{D}$ and $\mathcal{H}$, it's just going be w/ known data sets

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

D_g is $(x, \bar{y})$, and all the possible candidate fcs.

R_g is our h^*

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

yes, it's the best fc of all the candidate fcs from $\mathcal{H}$

(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 $\hat{y}^*$.

$\hat{y}^* = f(x^*)$

(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 bc the way Supervized learning works is we only ~~take~~ make our models from data that we have information on

(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.

$$g(x_1 \dots x_p) + \underbrace{\left(h^*(x_1, x_p) - g(x_1 \dots x_p)\right)}_{\text{estimation error}} + \underbrace{\left(f(x_1 \dots x_p) - h^*(x_1 \dots x_p)\right)}_{\text{misspecification error}} + \underbrace{\delta}_{\text{ignorance error}}$$

The error we get from estimating our x's

The errors we get when $\mathcal{H}$ is too broad

the error that will come w/ every model

(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.

M.E.: Redefine $\mathcal{H}$ to be more extensive

I.E. try to find more accurate x's

E.E. Create better methods of estimation

(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?

(w) [easy] What is a null model g_0? What data does it make use of? What data does it not make use of? It only contains the $\vec{y}$, it doesn't contain any t_j

(x) [easy] What is a parameter in $\mathcal{H}$?
a variable within these candidate fcs that the algo adjusts to better fit the data

(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?
g_0 = model [$\vec{y}$] bc there are no t_j to put in

(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?

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

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)

Feynman says we have to set the experiment, but our book says our data we need is already there

(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)

(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)

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

(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 kind of agree, just be it disagrees w/ experiment, there can be a lot of unknowns that haven't been tested so you can't objectively agree or disagree

(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? Models are never 100% right - that's why they're only models, but we can quickly tell when there's a bad (wrong) model compared to a good one

(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?

That there are always going to be some metrics out there that we haven't seen or measured, so just be the theory looks wrong, you can't say for certain that it is

(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)? If you can say ahead of time exactly how much love is enough and how much love is indulgence

(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?

bc these are psychological matters and they can't be defined so precisely

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.

when he explains it, it sounds simple. But to actually do what he says is extremely complicated. To push everyone out and become 2x more powerful than the next powerful country is much easier said than done

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