MATH 342W / 650.4 Spring 2021 Homework #1

Professor Adam Kapelner

Due 11:59PM Thursday, February 11, 2020 by email

(this document last updated 10:25am on Friday 5th February, 2021)

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: Meihe Liu

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?

most failure of prediction— stem from this false sense of confidence. Precise forecasts masquerade as accurate ones.

No.

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

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

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

(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.

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

(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.

they didn't use enough proxies for their model

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

risk is something you can put a price on

uncertainty is risk that is hard to measure.

(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.

(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.

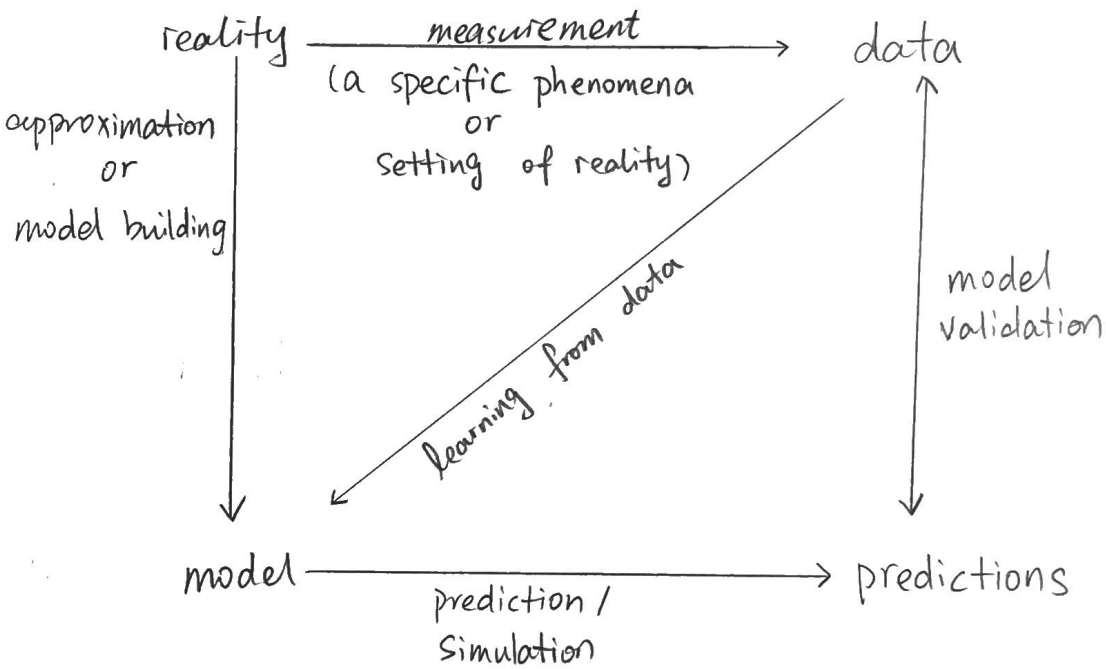
bias could describe accuracy

and variance could describe precision.

Problem 2

Below are some questions about the theory of modeling.

(a) [easy] Redraw the illustration from lecture one except do not use the Earth and a table-top globe. 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?

natural result of phenomenon / a being measured

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

by learning from data, to create a model and predictions are the outcome of the model in certain settings.

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

By definition, "models" are approximations which are not reality.

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

Because "Some models" are good enough to be used for a practical purpose

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

- good models would be unambiguous models. And the predictions we get by giving the model a certain setting would be very well related to the reality.
- bad models would be ambiguous models. And the predictions of the model wouldn't be so reliable.

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, because there are one input "an apple a day" and one output "good health", the inputs and outputs are numerical

∴ This is a mathematical model.

(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?

keeps the doctor away (good health)

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

I think

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

Let a = an apple a day

$x_1 = [a]$

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

Let g = keeps the doctor away

$y = [g]$

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

phenomenon, response, outcome, endpoint, dependent variable...

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

they are causal inputs

(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.

features, characteristics, attributes, variables, independent variables, covariates, regressors.

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

$y = t(x_1)$

it is approximate.

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

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

Because the approach we use is "learning from data" which is an "empirical approach".

(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. $\mathbb{D} = \{\langle \vec{x}_1, y_1 \rangle, \langle \vec{x}_2, y_2 \rangle, \cdots, \langle \vec{x}_n, y_n \rangle\}$

x_i's would be yes or no that the person eats an apple a day

and y_i's would be yes or no that the person is in good health.

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

- $\mathcal{H}$ is a set of candidate functions that approximate f.
- $\mathcal{A}$ is the algorithm that takes in $\mathbb{D}$ and $\mathcal{H}$ and returns a function g that approximate f.

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

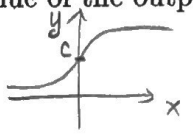
the domain and range of g should be the same as the domain and range of the $h \in \mathcal{H}$.

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

Yes! by definition of g, it is an output of the Algorithm $\mathcal{A}$ that takes in $\mathcal{H}$.

(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 $\hat{y}^*$.

$y^* = (x^* - c)^{1/3}$



(s) [harder] In lecture I left out the definition of f. It 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! $\mathcal{H}$ contains simply functions that can be fit with $\mathcal{A}$. However, f is very complicated and unknown. h could be very closely approximated to f but $f \notin \mathcal{H}$.

(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.

$$y = g(\vec{x}) + e$$

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

$\mathcal{E}$

ignorance error: difference between the reality and the model function f

model misspecification error: difference between the approximated function h^* and f

estimation error: difference between h^* and g while estimating data.

full error: the difference between predicted and observed.

(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.

- To reduce model misspecification error:

  Expand the set of candidate function $\mathcal{H}$ to be more complicated and thus more expressive of complex relationships.

- To reduce estimation error:

  Increase sample size n (more historical examples).

(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?

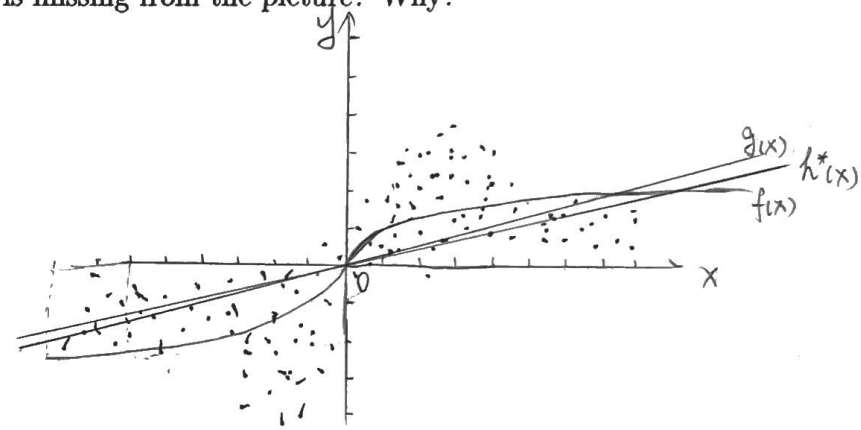
$f(x) = x^{1/3}$

$\mathcal{H} = \{\text{all linear models}\}$

$h^*(x) = \frac{1}{4}x$

$g = \mathcal{A}(\mathbb{D}, \mathcal{H})$

$= \frac{5}{16}x$



estimation error: $h^*(x) - g(x)$

model misspecification error: $f(x) - h^*(x)$

ignorance error can't be found from picture because we don't have data.

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

$g_0 = \text{mode}[\vec{y}]$

g_0 make use of phenomenon (y's) but not make use of proxies (x's)

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

θ

(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 average value

(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?

$\mathcal{H} = \{ \mathbb{1}_{x_i \geq a + bx} : a \in \mathbb{R}, b \in \mathbb{R} \}$

a is the intercept and b is the slope for the lines.

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

$y = 3 + 2x$

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)

Scientific method : Guesses → Compute consequences → Computation result vs. Experimental 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)

predictions

(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)

data

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

model validation

(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?

(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?

(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?

(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)?

(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?