

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 L^AT_EX. Links to installing L^AT_EX and program for compiling L^AT_EX 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 L^AT_EX, 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: _____

Marc Romero

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 interchangeable today?

Yes as they are calling the future
through past data.

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

very P/c of bias : small #
of true hypothesis, most studies
are likely to return false.

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

Ability to change
(couldn't find it)

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

How we process it.

- (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. $t, f, g, h^*, \delta, \epsilon, t, x_1, \dots, x_n, \delta, D, \mathcal{H}, \mathcal{A}, \mathcal{X}, \mathcal{Y}, X, y, n, p, x_1, \dots, x_p, x_1, \dots, x_n$, etc.

$$D = \left(\langle \vec{x}_1, y_1 \rangle, \langle \vec{x}_2, y_2 \rangle, \dots, \langle \vec{x}_n, y_n \rangle \right)$$

- (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 occurred? Answer using concepts from class.

For defaulting, all will default
or none will go binary {0,1}

Do not grade on a curve?

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

Risk is knowing the value of failures while
uncertainty is not knowing what can occur.
↓
Pot
a
prob
↓
lost to
measure

- (i) [difficult] How does Silver define *out of sample*? Answer using notation from class i.e. $t, f, g, h^*, \delta, \epsilon, z_1, \dots, z_t, \mathbb{D}, \mathcal{H}, \mathcal{A}, \mathcal{X}, \mathcal{Y}, X, y, n, p, x_1, \dots, x_p, x_1, \dots, x_n$, 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.

Out of sample - \rightarrow examples that
do not involve same setting

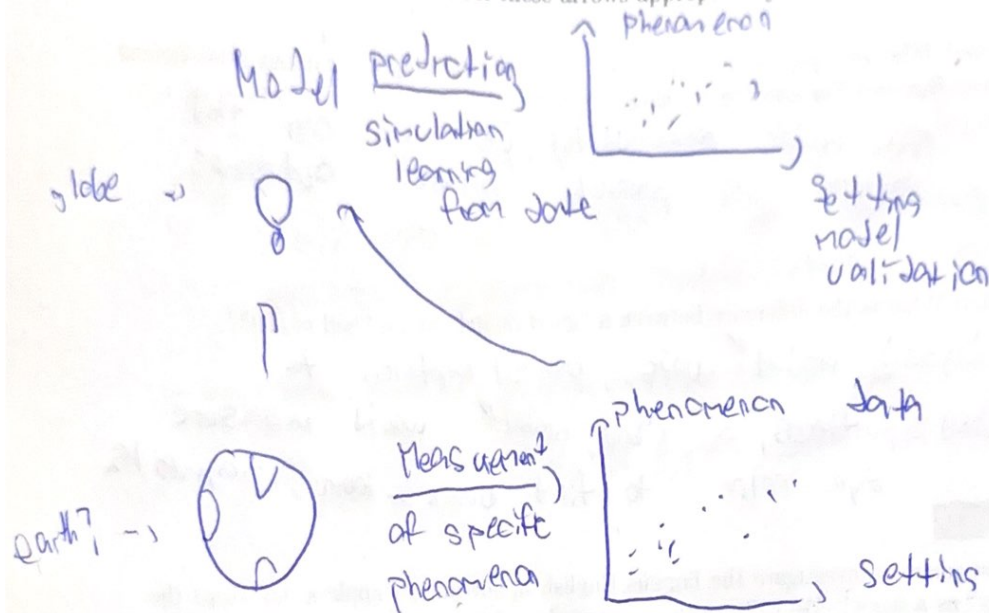
- (j) [harder] Look up *bias* and *variance* online or in a statistics textbook. Connect these concepts to Silver's terms *accuracy* and *precision*. This is another example of Silver using non-standard terminology.

bias : variance is estimating a assumptions
in order to make a better approximation

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?

Data is a model of reality.

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

Simulation learning from data

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

All models are wrong b/c they can't predict reality due to outside circumstances.

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

Some models are useful b/c we can find patterns to possibly predict outcomes.

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

A "good model" uses useful metrics to find patterns, a "bad model" won't measure eye color to find bank loan payments.

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 b/c we could measure apples ate to physical health.

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

Apples eaten by a person. ~~and~~

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

Physical health

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

It is hard to correlate bc physical health can be affected by numerous variables, which an apple cannot stop.

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

$x_1 = \text{QOL}$ metric Apples ate ϕ over a year

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

$y = \text{QOL}$ metric

(g) [easy] What is y mathematically?

Continuous metric

$\phi(16; 112) \rightarrow$

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

z_1, \dots, z_t is the true drivers of the phenomenon

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

$$p=1$$

- (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.

 \rightarrow all real numbers

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

$$D = (X, \vec{y}) \rightarrow \text{approximate}$$

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

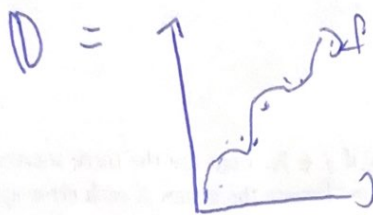
Algorithm taking D as inputs ; locating the 'best' element of H to fit the lowest possible error.

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

we find the function w/

lowest error given the inputs.

(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.



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

\mathcal{H} = Gives lowest error function

\mathcal{A} = ~~Reduce~~ Reduce misclassification error

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

Dom g in = All models

range = \mathbb{R}

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

Yes b/c $g = \mathcal{A}(\mathbb{D}, \mathcal{H})$
thus, takes all
models of \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}^* .

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

- (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?

- (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.

- (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.

- (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?

$$g_0 = \text{model}[\vec{y}] \in \{0, 1\}$$

we only see \vec{y} , : does not use x 's.

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

$$\mathcal{H} = \left\{ \mathbb{1}_{\vec{w} \cdot \vec{x} \geq 0} : \vec{w} \in \mathbb{R}^{p+1} \right\}$$

- (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?

$$\{0, 1\} \text{ s.t. } \mathbb{1}$$

- (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?

$$\mathcal{H} = \{ \mathbb{I}_{x \geq \theta} : \theta \in \mathcal{X} \}$$

$$\text{parameter } \{0, 1\}$$

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

$$g(x) = \mathbb{I}_{x \geq \theta_g}$$

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)

we guess w/ ~~data~~ scientific method (I guess
he means hypothesis)

- (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)

Phenomenon

- (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)

Model

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

Measurement

- (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 agree b/c
models are said to be wrong but
we would add that they are useful

- (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?

We cannot replicate reality but
we can prove certain theories
aren't accurate.

- (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?

We cannot prove it wrong b/c
we don't know what metrics
a vague theory is using and
cannot disprove it.

- (h) [difficult] He then he continues with an example from psychology. Remember 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)?

we need exact metrics to prove
such 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?

There is no way to
measure "love"

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.

Complicated b/c we don't know what
life will be like a time from now

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

China prefers realism, while the U.S.
tends to divert away from realism.