

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

Mohammed Hasan

Sunday 11th 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 interchangeably today?

The terms *predict* and *forecast* are used interchangeably today despite there being clear differences between the two during shakespearean times

- *forecast* : means to plan under conditions of uncertainty
- *predict* : is described as "what the soothsayer told you" meaning it is certain

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

John P. Ioannidis's findings detailed in the paper titled "why most published Research Findings are False" says that a significant portion of positive findings in biomedical research journals are not able to be applied to real-world settings. The implications of his findings are that it questions the reliability of the research. Bayer laboratories confirmed his hypothesis.

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

Human beings most powerful defense is their ability to recognize patterns.

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

Our human ability to process information is getting outpaced by the information that is 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. $t, f, g, h^*, \delta, \epsilon, \tau, z_1, \dots, z_t, \delta, \mathbb{D}, \mathcal{H}, \mathcal{A}, \mathcal{X}, \mathcal{Y}, X, y, n, p, x_1, \dots, x_p, x_1, \dots, x_n$, etc.

I believe the objective truth is: $y = t(z_1, \dots, z_t)$ where y is the response/dependent variable and the z 's are the proximal causes which the unknown function, t , uses to obtain.

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

Popper's definition of science relies on the idea of falsifiable. For him, a hypothesis is only scientific when it can be proven through real-world experimentation.

- (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 models that the rating agencies used were flawed and the agencies were relying on them heavily. pg38 "forecasting models were full of faulty assumptions and false confidence about the risk."

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

- *Risk* : can be evaluated numerically such as a price.
- *Uncertainty* : is harder to measure since it can not be defined numerically.

- (i) [difficult] How does Silver define *out of sample*? Answer using notation from class i.e. $t, f, g, h^*, \delta, \epsilon, z_1, \dots, z_t, \mathbb{D}, \mathcal{H}, \mathcal{A}, \mathcal{X}, \mathcal{Y}, X, y, n, p, x_{\cdot 1}, \dots, x_{\cdot p}, x_1, \dots, x_n$, 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.

Silver defines *out of sample* by giving an example of driving under the influence. Your problem is *out of sample* when you are using an X that is irrelevant. Your car trips driven sober should not be counted if you are driving drunk. This would be your first time driving drunk.

- (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 very similar to *accuracy* and *precision*. *accuracy* is the ability to land closer to the "Truth" which is the inversion of *bias*

High accuracy = Low Bias

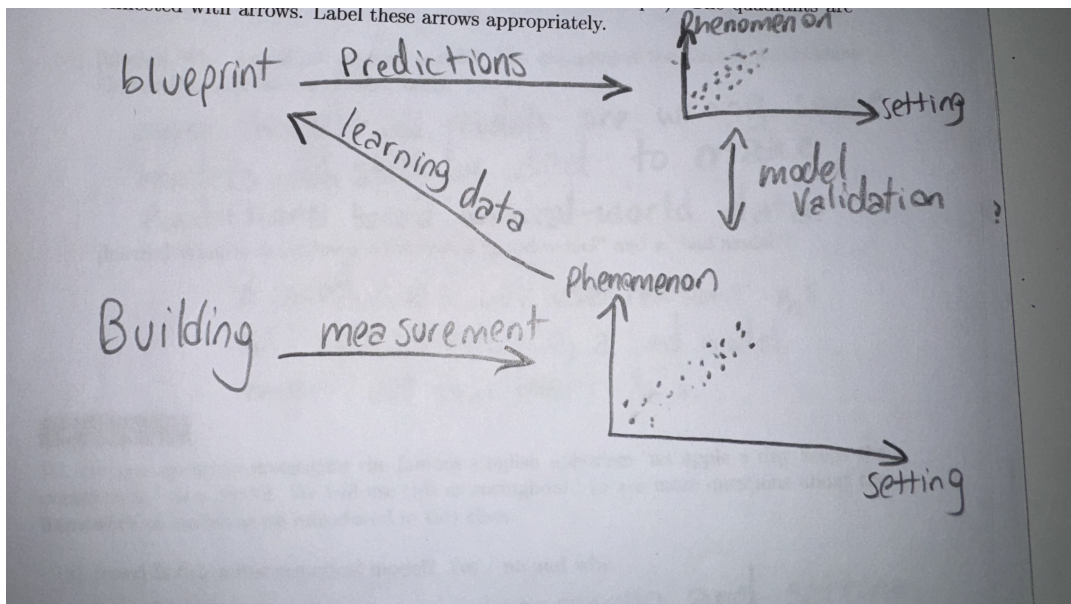
precision is the range of the cluster. If the cluster is further apart it has Low Precision

High Precision = Low Variance

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, for this class, pertains to results of events that already occurred that can be used to predict the outcome of another event.

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

According to our illustration, *Predictions* are expected calculations which we derive by interpreting the real world phenomenon.

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

“all models wrong” because they are not fully accurate and precise.

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

Even though all models are wrong, some models can still be used to make predictions based on real-world data.

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

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, we have a phenomenon and setting, we can also used metrics to measure them.

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

1 apple

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

Doctor away

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

I believe, although it encourages a healthy lifestyle, when tested, the model only accounts for 1 of the many phenomenon required to satisfy the setting.

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

x_1 : mean number of apples eaten yearly

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

y : mean number of times doctor visited yearly

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

\mathcal{Y} is a scalar that we will use to express a phenomenon.

- (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 in this phenomenon are the true drivers.
ate apples daily : (0,1)

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

p = number of \mathbf{x} 's
 $p = 1$ since only x_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.

x_1 numerically represents average apples eaten yearly

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

y and x_1 is approximate since the equation of x_1, \dots, x_n contains δ to account for the errors.

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

supervised learning is where data sets are used to predict outcomes. The more relevant input data is given, the closer the predictions will be to actual outcomes.

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

supervised learning is an empirical solution because it relies on data to be inputted for it to make accurate predictions.

- (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.

$|D| = \langle x, \vec{y} \rangle$
 $|D|$ would have a pair of inputs

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

\mathcal{H} is a set of candidate functions, \mathcal{A} is the algorithm which takes $|D|$ and \mathcal{H} .

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

Domain of g is all possible inputs and range is all possible outputs which are \mathcal{H}

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

$g \in \mathcal{H}$ because after \mathcal{A} completes, a g is picked 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}^* .

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

- (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?

yes, $f \in \mathcal{H}$ because \mathcal{H} is large enough to approximate f .

- (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.

$y = g(x_1, \dots, x_p) + (h^*(x_1, \dots, x_p) - g(x_1, \dots, x_p) + (f(x_1, \dots, x_p) - h^*(x_1, \dots, x_p)) + \delta$
where

$(h^*(x_1, \dots, x_p) - g(x_1, \dots, x_p)) =$ estimation error: the difference between predicted values and true values

$(f(x_1, \dots, x_p) - h^*(x_1, \dots, x_p)) =$ Miss-specification error: when model does not show true variable relationship

$\delta =$ ignorance error: factors simply not accounted for.

- (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.

- reduce estimation error: inverse n (add more data)
- reduce misspecification error: according to $ME = 1/n$ TE adding more n 's reduces error
- reduce ignorance error: adding more relevant x_n 's

- (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 is the null model used to predict \hat{y} since we do not have any x 's.
it makes use of $\text{Mode}[\hat{y}] \in 0, 1$ // it does not make use of x 's

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

$$\mathcal{H} = \{I_x : \theta \in X\}$$

- (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?

$$g_0 = \text{Mode}[\hat{y}] \in 0, 1$$

- (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?

$$\mathcal{H} = I_x \geq 0 : \theta \in X$$

Parameters: $\theta \in X$

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

In the example of credit score:

$$\begin{cases} 1 & \text{if } X \geq 650 \\ 0 & \text{if } X < 650 \end{cases}$$

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)

Starts with a guess

- (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)

Measurement

- (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?

agree, if the computation disagrees with the experiment, the computation should be reevaluated to agree with the experiment.

- (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?

This means our models will not ever be fully accurate but we can tell when they are wrong.

- (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 should keep our models and metrics less vague so that if our models are wrong, we should know.

- (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)?

The example was that he was not loved enough. Even though he was overindulged. meaning it was too vague enough to be on either side. Remedy was to ask metrics. How much love is not enough? how much is too much love?

- (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?

Because it was a psychological matter and "things couldn’t be defined so precisely"