

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

Loyd Flores

Due 11:59PM February 11, 2024 by email

(this document last updated 9:02pm on Sunday 11th 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](https://www.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: Loyd Flores

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?

Answer:

As Seraydarian defines, "Prediction involves estimating an outcome with a high level of certainty, usually based on historical data and statistical modeling. On the other hand, a forecast involves projecting future developments but with a certain level of uncertainty due to external factors that may impact the outcome." This helps us understand how Silver defines prediction. It is "what the soothsayer told you." Soothsayer being the data that tells us the story on how we can make predictions. Forecast on the other hand is defined as "implied planning under conditions of uncertainty. It suggested having prudence, wisdom, and industriousness." Which aligns with Cassius' idea that a man is a master of his own fate as forecasting utilizes the user's foresight to a greater extent. Although these differences are clearly defined, in contemporary usage, the two terms are often used interchangeably, a point also noted in Silver's book.

Citation: Seraydarian, L. (2023, August 11). Prediction vs forecasting. Plat.AI. Retrieved from <https://plat.ai/blog/difference-between-prediction-and-forecast/>

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

Answer:

John P. Ioannidis released a paper titled "Why Most Published Research Findings Are False." Bayer Laboratories confirmed John's findings, as they were unable to replicate about two-thirds of the positive findings claimed in medical journals when they attempted the experiments themselves. In my interpretation, John's findings imply that the experiments in these published journals may have had controlled variables which could not be replicated. We can conclude that the findings were false due to the irreproducible nature of these findings, which means they are not true in all cases or when other external variables are introduced.

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

Answer:

Silver asserts that our most potent defense lies in our wits. Human minds are more agile than those of any other species, uniquely equipped to discern patterns and rapidly respond to both opportunities and threats. This aligns with our class discussions, where we emphasized humans' capacity to learn and recognize patterns through data

analysis. Furthermore, humans excel in assessing situations, as displayed by Jefferson's quote, "Early to bed, early to rise makes a man healthy, wealthy, and wise." This proverb presents our ability to evaluate and improve our circumstances proactively. In the modern era, we have even extended this capability through technology, developing methodologies like machine learning to process and interpret data at unprecedented rates, thereby augmenting our inherent strengths.

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

Answer:

Since the development of the printing press in the 1400s, there has been a rapid increase in the production of information, as evidenced by the proliferation of books and other forms of consumable text. This acceleration is also reflected in the timeline of human development: it took 300 years to transition into the industrial age, but only 200 years after that for the internet to become a dominant force in our lives. The technological advancement gap is narrowing further, with AI emerging just 20 years after the internet's inception. However, as Silver notes in his book, what is not keeping pace with this rapid increase of information is the human capacity to process and digest it. The surge in data availability means there is also a corresponding rise in false or misleading information, creating a challenge for humans who must now sift through this vast expanse of data without adequate time for proper analysis and understanding.

- (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. $t, f, g, h^*, \delta, \epsilon, t, z_1, \dots, z_t, \delta, \mathbb{D}, \mathcal{H}, \mathcal{A}, \mathcal{X}, \mathcal{Y}, X, y, n, p, x_1, \dots, x_p, \dots, x_n$, etc.

Answer:

In class, we discussed the concept of 'objective truth,' denoted as 'f'. It is the closest approximation we have to 'y', which are real world results, and ideally fits our data. However, the objective truth, or 'f', is unattainable due to its complexity. The closest we can get to 'f' is with 'h*', a function within 'H', our set of functions or hypothesis set, which has the lowest error. The function 'h' is selected by 'A', our learning algorithm. With everything previously stated, we can develop 'g', our model that takes 'x' (which could be values from 'X', our input space) to produce an approximation of 'y'. The model 'g' is formulated by inputting 'D', our data or a sample of our data, and 'H' into 'A'. This process produces the best combination of these components to come up with our model, 'g'.

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

Answer:

For Karl Popper a philosopher of science states that science is composed of hypothesis that could be tested in the real world by means of prediction. In a nutshell if we are unable to test a hypothesis in the real world then it is not scientific.

- (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.

Answer:

Referencing concepts from our class the rating agencies predicted only .12% due to the availability of their data or in our lectures terms 'D'. Their model was based off of previous events and they failed to account for what ended up happening. The error they had was really due to ignorance because the data they needed was not available. One can argue that they could've based off of the housing market crash in Japan that happened which was in recent history during those times.

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

Answer:

Silver defines 'risk' as something we can quantify or put a price on. We understand the likelihood of such risks occurring and the potential effects if we experience the consequences of those risks. On the other hand, 'uncertainty' refers to risks that are difficult to measure. With uncertainty, one can never fully comprehend the entire context. The best one can achieve is a vague understanding of the situation, but never a complete picture.

- (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.

Answer:

Silver defines out of sample as something that had never been seen before. He gave the example of the rapid increase in U.S. housing that preceded the collapse. In our

lectures context there were data points 'x' that were not in 'X', our input space. The model that we have was only familiar with data points in our 'D' and was performing well because it had very low 'error'. The problem is a point 'x' that was out of sample and the model was only able to predict based on the available data in 'D' making the prediction wrong because of a large error due to ignorance.

- (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.

Answer:

In a statistical context, bias is the error that occurs when we approximate a real-world problem in simplified terms, despite its complexity. Bias measures how far off a model's predictions are from the actual value. High bias may indicate that a model is oversimplified and fails to consider all variables or relationships. Variance, on the other hand, is defined as the model's sensitivity to fluctuations or noise in the dataset. It measures how predictions for a point might vary between different realizations of the model or the extent of the spread between predictions. A model with high variance is so complex that it captures even the noise within the data patterns, performing well with training data but failing with unseen data. Connecting this to Silver's definitions: bias relates to the accuracy of the model, meaning that if the model makes errors, it will consistently do so in the same direction. It may result in our model consistently hitting the same spot, but not the target. Variance is related to precision. As previously defined, variance is the spread of the data; in Silver's terms, it addresses how far off our shots are. Low precision equates to high variance. If our shots are all close together, we have high precision or low variance. This can mislead the engineer because high precision does not necessarily equate to quality.

Citation:

Bias (Statistics) (2024/02/10). In Wikipedia. [https://en.wikipedia.org/wiki/Bias_\(statistics\)](https://en.wikipedia.org/wiki/Bias_(statistics))

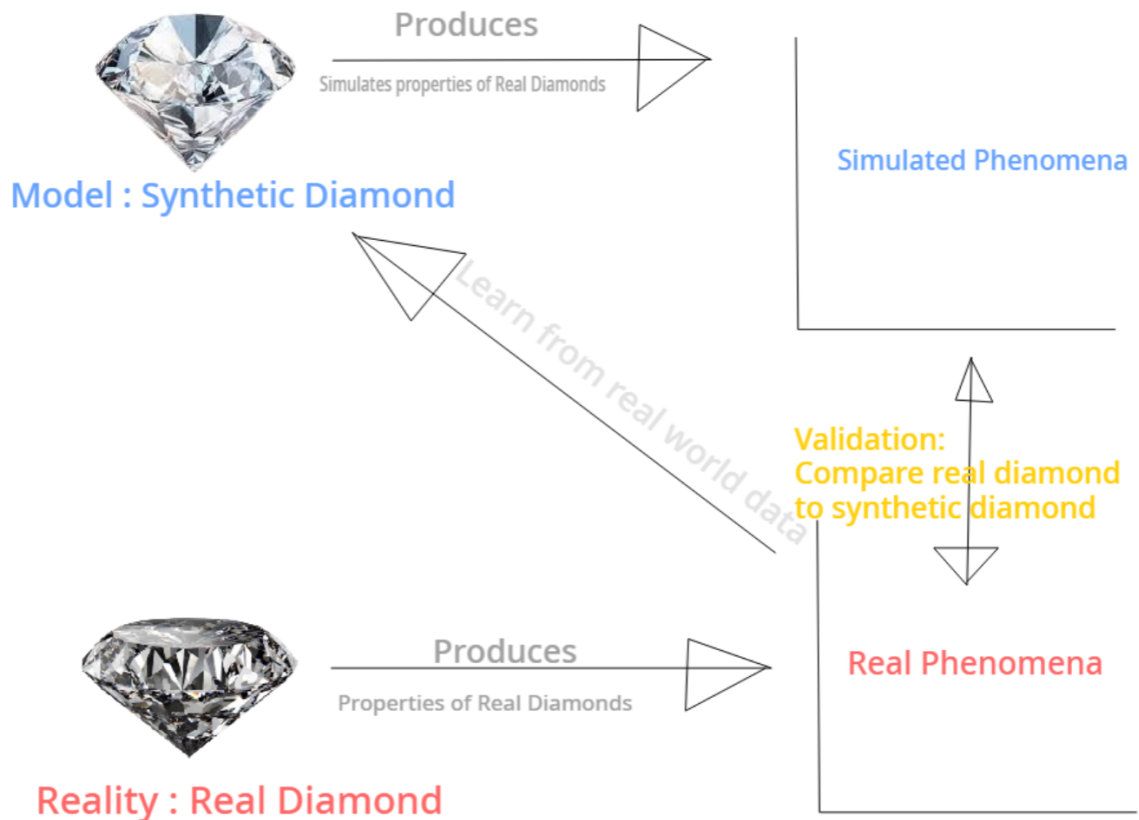
Hayes, A. (2023, December 19). *What Is Variance in Statistics? Definition, Formula, and Example*. Retrieved from: <https://www.investopedia.com/terms/v/variance.asp>

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.

Image :



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

Answer:

Pursuant to the previous question data is defined to be the properties of something real. Referring to the previous example it could be the size, weight, hardness, and other properties of a diamond that could produce real phenomena as such as how it is created or how it could withstand immense pressure. In the context of our class the diamonds' properties could be our 'D'

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

Answer:

Pursuant to the previous question prediction in our class is defined as the actualization of our hypothesis on the available data that we have. Using the illustration presented the prediction is a synthetic diamond. We came up with this conclusion using real properties of a diamond. Calculations are made to simulate the diamond which could be our 'h'. We can then validate our synthetic diamond to the real one by comparing

their properties.

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

Answer:

George Box says "All models are wrong but some could be useful". This is true because a model is just an approximation of reality. Models will never be the real thing. Some models are useful because it can help us navigate reality but as previously stated can never be the absolute truth.

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

Some models are useful because they are really fine tuned. They can produce really close results to reality sometimes the difference can be so minute you can assume it is reality, keyword ‘assume’. Models are still models meaning they are only approximations and we can only assume. It is not the absolute truth therefore it can be useful to guide us.

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

Answer:

In my understanding a good model accurately predicts the outcome desired and generalizes well to unseen data while a bad model fails to make accurate predictions beyond the data it was trained on, sometimes failing to even make accurate predictions on training data.

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.

Answer:

In my opinion It is not a mathematical model. It is merely a proverb. A mathematical model is a little more complex than a proverb. There has to be a quantifiable relationship. Also I don't see what kind of equation we can produce out of this. Although if I had to make this a mathematical model, I could attempt to learn the relationship of consumed apples to number of doctor visits both in an annual time frame.

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

Answer:

The inputs for this could be the following ‘x1’: # of apples consumed daily, ‘x2’: # of

annual doctor's visits. We could optionally add other measures of health such as the amount of times you caught a cold or got sick, etc ...

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

Answer:

The output of our model could either predict the number of visits annually, instead of it being an input it could be our output. We could also produce a model that would accept x being the number of apples consumed and it would determine whether or not it will reduce your # of visits or if it improved via a binary response y/n where it would learn through data collected.

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

Answer:

I think this model will be bad. It would be as good as a random person predicting this. I don't think there is a correlation because yes an apple is nutritious but it isn't the sole determinant to whether or not it can keep you away from the doctor. There are other factors on why you would visit the doctor that is not health related. I don't see an empirical solution to this problem just by the number of apples consumed.

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

Answer:

As I previously stated above ' x_1 ' could be the total number of apples consumed annually or it could be average apples consumed per month. The consumption rate could be scaled up or down depending on how the Data scientist will prefer their units.

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

Answer :

' y ' could possibly a binary classification of y/n that determines if you visit a doctor in a specific time frame, or if your medical bills increased or decreased given a time frame. It could also become a regression model where it tries to predict medical bills or total number of medical visits.

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

Answer:

$\mathcal{Y} = g(x)$, where g is our model and ' x ' in ' X ' which is our input space.

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

Answer:

' z 's are what contribute to our prediction ' y '. ' z 's could be a wide array of things that could potentially affect our result. In this phenomenon the ' z 's could be the number

of apples consumed for example. This could affect whether or not we go to the doctor or how much we spend on the doctor.

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

Answer:

I defined my 'x1' to be average number of apples consumed in a month. With that being said this in itself will not predict p or even contribute that's why I could say that $p = .5$. We have a 50-50 chance to getting it wrong because its no better than a random guess.

- (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.

Answer:

$X = \{x \in \mathbb{R} \mid x \geq 0\}$ We give this formula to make x quantifiable. The least we can have is 0, we cant have a negative consumption of apples.

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

Answer:

$y \approx f(x)$ We can only approximate y and it will never be equals.

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

Answer:

Supervised learning is when we have a target label available in our data. We can use this target label to verify our predictions to see whether we are right or wrong. This is where the term supervised comes from because we are using the target to guide us.

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

Answer:

Supervised learning is considered to be empirical because it is built solely on available data. It is data driven in a sense that it needs targeted data sets to be considered a supervised learning model. It will be great to have theoretical background but it requires the data more than anything else hence an empirical solution.

- (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.

Answer: $\mathbb{D} = \langle X_1 \dots X_n, Y_n \rangle$. $X_1 \dots X_n$ would be the column or features and it would range from 1 to n which would be the number of columns. Y would be the result or the prediction obtained using the available data.

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

Answer:

In the framework of machine learning, \mathcal{H} represents the hypothesis space, which is the set of all possible functions or models that could be used to fit our data D . This space encompasses every conceivable model.

On the other hand, \mathcal{A} is the learning algorithm. Its role is to search through the hypothesis space \mathcal{H} to find the most suitable model that effectively captures the relationship inherent in the data D . The algorithm does this by evaluating different models within \mathcal{H} based on their performance, often using a criterion such as minimizing error.

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

Answer:

The domain of this problem will be all possible input pairs in \mathbb{D} evaluated by all possible 'h's in \mathcal{H} . The range will be the dot product of the two or all possible outcomes from all the configurations of data points and evaluating functions.

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

Answer:

$g \in \mathcal{H}$ because \mathcal{A} selects the best performing h in \mathcal{H} . The g we end up with performed the best by attaining lowest error in the set of all available functions.

- (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}^* .

Answer:

$\hat{y}^* = g(x^*)$. Where the unseen data is taken by our model to spit out a prediction.

- (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?

Answer:

It is highly unlikely that f is in \mathcal{H} because f perfectly fits to our data. f is something that we are trying to attain. If anything f maybe too complex to even exist due to its perfect nature. \mathcal{H} is the set of available functions that we have sometimes it is over simplistic by being only linear models.

- (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.

Answer:

The entire equation is as follows $y = g(x_1, \dots, x_p) + (h^*(x_1, \dots, x_p) - g(x_1, \dots, x_p)) + (f(x_1, \dots, x_p) - h^*(x_1, \dots, x_p)) + \delta$. Where $g(x_1, \dots, x_p)$ is our model. Estimation error is basically when our model is unable to capture the relationship of our data points this could occur because either we have too few data or there is too much noise in the data we have. It is denoted by $(h^*(x_1, \dots, x_p) - g(x_1, \dots, x_p))$, we can

reduce estimation error by collecting more data so g converges to H^* . The next error is specification error which is caused by our model oversimplifying the relationships in our data we can solve this by utilizing more complex models. It is denoted by $(f(x_1, \dots, x_p) - h^*(x_1, \dots, x_p))$. The last type of error is δ which is denoted by δ or could be referred to as Ignorance error. This is basically the error allotted because we don't really know or have data about certain things, the best we have are proxies.

- (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.

Answer:

Estimation error could be reduced by adding more data. With more data our model will be more capable of identifying relationships between variables. Mis-specification error is caused by our model being unable to fit to our data maybe because it is too complex. This is solved by utilizing more complex functions that will be able to fit with the data. The last error is error due to ignorance which is the hardest one to compensate. The only way I can think of is by collecting domain-specific data or having an expert on the team to add more analytical judgement.

- (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?

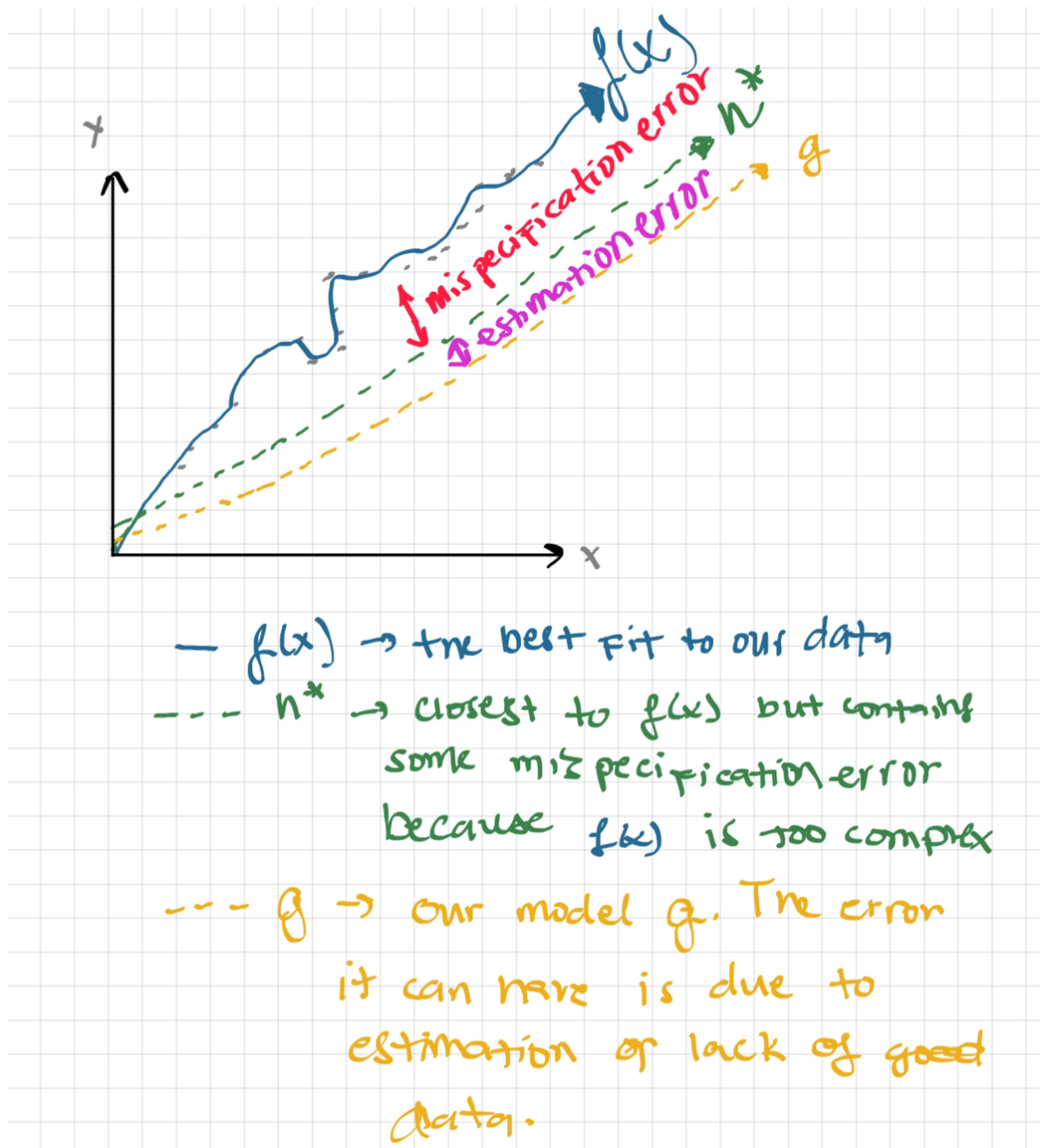


Image:

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

Answer:

g_0 or the null model is a baseline model only taking into consideration y and none of the other features in X . The purpose of this model is to benchmark the performance of other models that could be more complex. If a model fails to outperform the null model it means that said model is unable to capture the relationship in our data or the engineer probably did something wrong in the process.

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

Author:

In my understanding a parameter is a value that defines a particular model within the space. each h in \mathcal{H} could be characterized by a set of parameters which determines how a model behaves and fits data. For a regression model it could be the intercept of a line. Intercepts could define how h will interact with the data. For complex models like neural networks parameters could be weights of different variables which would be adjusted through iterations.

Citation: n/a (2024, February 11). "Hypothesis Parameters in Machine Learning". Retrieved from <https://chat.openai.com/c/dedf1ec2-b14d-4eac-8729-0f0bd4863f0c> (ChatGPT)

- (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?

Answer:

The null model g_0 is a baseline model that only considers y . It will predict the most frequent occurrence of a class in y . The accuracy then becomes the percentage of the appearance of that most frequently occurring class in our dataset.

- (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?

Answer:

\mathcal{H} would consist of all functions that could produce a binary output. The parameters in this context would be the threshold. Each h in \mathcal{H} could have different set threshold as they will all interact with the data differently. In summary \mathcal{H} will consist of different ways we can combine data and the parameter would be the threshold which would be how they each method of combination would handle data differently.

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

Answer:

For example $g = 2x+3$ and our threshold is 5. Our model takes x as input and will spit out a binary classification of yes/no. If $g(x) > 5$, then the output is yes. Else $g(x) < 5$, the output is no.

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)

Answer:

Feynman discusses that the scientific method involves. First we guess it, then we compute the consequences of the guess to see what it implies, then we compare the results to nature or experiment. If it disagrees it is wrong. Learning from data approaches this differently by analyzing existing data to find patterns before making any predictions which is opposite to Feynman's scientific method.

- (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)

Answer:

In the context of our lectures we used the term 'Validation' where we compare the phenomena of our simulation or model to the phenomena of real world data.

- (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)

Answer:

In the context of our class instead of experiment we use prediction or learn from data. The prediction is our experiment based off the patterns we learned by analyzing our data.

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

Answer:

Validation or where we compare our model's phenomena to the real world's phenomena.

- (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?

Answer:

Feynman's quote says that if your hypothesis doesn't align with the experiment there is no way around it, your hypothesis is just wrong, plain and simple. A data scientist may argue because the objective function we possess as of now is way too simple. The necessary model maybe more complex. Another concern could be the data, it may just mean we lack data. A data scientist may argue that a process is not just binary right or wrong, there maybe more underlying reasons as to why our model could be wrong now but could be corrected looking forward.

- (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?

Answer:

Feynman quotes that we only prove guesses wrong and never right. As long as a theory stand it's not wrong but also is not right, it just stands. This happens due to the lack of knowledge we have in status quo. Something may be understood the way it is now but in the future more complex systems and analyses could exist that could disprove a theory that stands today. In the context of mathematical models it is pretty similar. There are some models that are unable to fit towards data that is present due to its complexity. For example before they only had the Perceptron Algorithm that is able to find the best model that could split the data into 2 sections. Throughout time more complex models were developed and the perceptron is barely used in present day.

- (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?

Answer:

If a theory is not clear and specific, it's hard to prove it wrong because it doesn't clearly state what it's about. It uses unclear standards that are not measured correctly. There might be parts of the theory that don't really show how they are connected to what you're trying to prove, making you think the theory could be right just because it's not specific enough. Proving such a theory wrong is tough because its vagueness can make it seem right, even if it isn't. It would also be hard to disprove such vague theory because it will be difficult to test against it due to its vague nature.

- (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)?

Answer:

Feynman states that "It would be possible to state ahead of time if how much is not enough and how much is overindulgent." This means that if we were to quantify the available parameters we have better by having thresholds it would make our hypothesis testable. He says that if we quantify our problem better then an explanation can come out of it.

- (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?

Answer:

In my interpretation this means that if your hypothesis is vaguely defined or if the variables in the hypothesis is not quantified properly despite whatever answer you get it is not correct. You can't claim to know anything when your hypothesis can't even be tested or proven right or wrong. If you can't even define your problem, you won't be able to answer it.

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.

Answer:

John Mearsheimer's model of great power politics is both simple and complicated. It depends on the perspective from which it is viewed. It could become simple because there is a clear theoretical framework. It is based on five basic assumptions about the nature of international relations, making it straightforward. It is also consistent with realism. He aligns the model with realist perspective. With the previous points mentioned the model then becomes predictable which makes it simple. It can also become complicated because when applied to diverse scenarios adding the complexity of the real world and international relations the model then becomes really complicated. The diversity of states, contexts, history, and etc.. add even more layers of complexity. We also then have to understand the intentions of other states'. This is a problem because we have to sort situations out where everyone wins which most of the time is impossible. We also have to add inconsideration the natural of power politics. The shifting balance of power makes the context a little more dynamic. Solutions or negotiations that could happen now may not be able to push through tomorrow. In summary the model in itself is simple and straightforward it gets difficult when applying this to the real world where there are a multitude of variables that could change the outcome.

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

Answer:

John Mearsheimer states that most theories can never get everything right and all theories have limits because theories are simplification of real world scenarios. He even states that the best models can get things right 75% of the time and wrong 25% of the time. In the context of our class models behave in the same way. The best models today are the best of status quo but could be developed as years go by. Our models will always have limitations and can never be the absolute truth mainly because it is only an approximation of reality. Models can also fall short to other things such as oversimplified models that attempt to understand very complex data. There is also a big portion of error that models obtain due to the fact that we don't really know the entire context of the real world. There is just error due to ignorance. As we develop our models like John states we hope to further develop our models more and more as time goes by.