MATH 390.4 / 650.2 Spring 2020 Homework #1

Tyron Samaroo

Wednesday 12$^{\text{th}}$ February, 2020

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?

During Shakespeare's time *predict* and *forecast* had two different meanings. A prediction is something that a fortune teller or a soothsayer would tell you. A forecast is implied planning under certain conditions of uncertainty. Today *predict* and *forecast* are both used interchangeably.

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

John P. Ioannidis publish a paper "Why Most Published Research Findings Are False." He concluded in his paper that predictions of medical hypotheses in medical experiments were most likely to fail if applied in the real world. This is trying to imply that having full dependency on predictions can be catastrophic on society especially when they end up being wrong.

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

Humans being's most powerful defense according to Silver is our need to detect patterns. We try to predict the pattern of future events of a phenomenon based on our own examinations.

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

Information is increasing rapidly but the amount of useful information is not. Silver consider the surplus of information as noise and the useful information that's there as the signal which is the truth that we want.

(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.

We describe the objective truth as a phenomenon y is equal to a true function t with causal inputs or "*features*" $z_{1\cdot}, \ldots, z_n$

$$y = t(z_{1\cdot}, \ldots, z_n)$$

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

Karl Popper's definition of science is that a hypothesis was not scientific unless it was falsifiable. This means that if the negation of a hypothesis can also be proven then it would 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.

The rating agencies were wrong with their prediction because they did not have any historical data so instead they made faulty assumptions that ended up being catastrophic.

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

According to Silver *risk* is define as something you can put a value or price on. *Uncertainty* is define as risk that is hard to measure there's no good way to measure this. An example that sums that up provided by silver would be risk is like the greases the wheels of a free-market economy while uncertainty grinds them to a halt.

(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.

Silver explains out of sample with an example of a good driver. A good driver has data that he's a good driver by having 20,000 trips trips in which two resulted in a car accident but still has a good 19,998 trips where he got home safe.The driver got drunk and assumes that since he was able to make majority of his trips home with a high probability then theirs a high probability that he would make it home drunk.The problem is that the driver has no data on being drunk so his sample size isnt 20,000 but 0 and. this is what Silver consider out of sample.

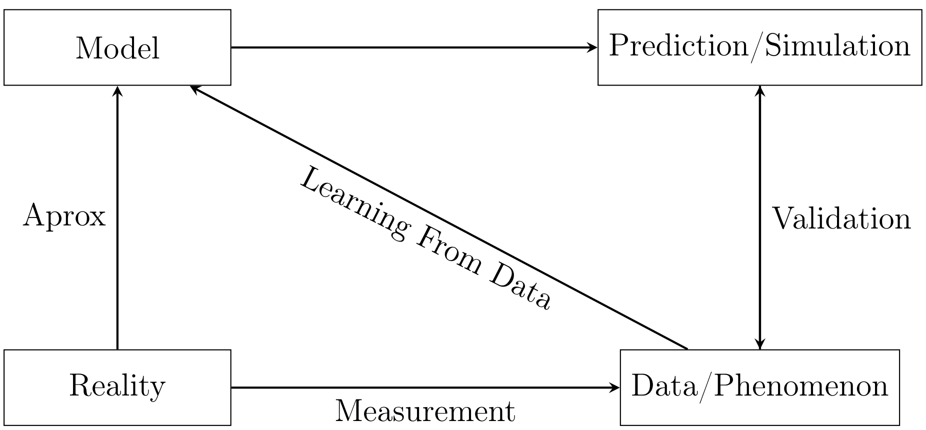
(j) [harder] Look up *bias* and *variance* online or in a statistics textbook. Connect these concepts to Silver's terms *accuracy* and *precision*. This is another example of Silver using non-standard terminology.

Bias tries to oversimplifies things and miss the important parts leading to high errors. Variance tries to over over compensate and fail to generalize the data, it performs well on testing set but not on real data. Silver term precision seems to be connect to variance and accuracy to bias.

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?

In this class we define data as being natural results of measuring a phenomenon.

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

In class we define prediction to be a phenomenon under examination that can modeled.

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

All models are wrong because given a phenomenon y we are able to understand it only if we know the true function t with casual inputs $(z_1., \ldots, z_n)$ but we do not.

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

Some models are useful because we can try to approximate the casual inputs $(z_1., \ldots, z_n)$ which we can call $(x_1., \ldots, x_n)$. In doing so we can try to approximate the phenomenon the best we can.

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

A "good model" is a model that can make useful predictions given a set of inputs. A "bad model" on the other hand given the same set of inputs result in wrong or not useful predictions.

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 this is a mathematical model because it is saying given an input apple a day you will have an output which is keeping the doctor away.

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

This model input would just be the apple.

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

The output would be whether or not the the doctor was kept away.

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

This model is bad because based on how it is phrased we are restricted with one input to make a prediction of the given phenomenon.

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

A metric for gauging the main input can be represented as x_1 which can be whether or not a person ate an apple

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

A metric for gauging the main output can be represented as y which would be whether or not the person did not have a visit to the doctor

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

$\mathcal{Y}$ mathematically would be the phenomenon that we are trying predict when given a true function t with causal inputs $(z_1, \ldots, z_t)$

(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 $z_1, \ldots, z_t$ are the true casual inputs that can be passed in the function t to predict the phenomenon.

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

If we let p denote the number of features or attribute, so that we have a vector $x_i := [x_{i1}, \ldots, x_{ip}]$

(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.

$\mathcal{X}$ mathematically is the input space such that $x_i := [x_{i1}, \ldots, x_{ip}] \in \mathcal{X}$

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

We term x_1 to be one of the features which will be a independent variable, but x_1 is just an approximation for y.

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

Supervised learning is where we have historical data which can be consider training data that is labeled which we can learn from.

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

Supervised learning is an *empirical solution* because it is using historical data to make a prediction where analytical solution is given from an integral or derivative.

(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.

$\mathbb{D} = [\mathcal{X}, y]$ $\mathbb{D}$ is our data that has historical examples with label responses.

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

$\mathcal{H}$ is a set of candidate function h that can approximate f. $\mathcal{A}$ an algorithm that takes $\mathcal{H}$ and $\mathbb{D}$ and provides g $\in \mathcal{H}$ as the best approximation of f which would be h*

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

The domain should be the inputs from data D and the set of H. Its range should be the output of the approximation of the phenomenon y

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

$g \in \mathcal{H}$ because we defined an algorithm A which produces g when given data $\mathbb{D}$ and 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 $\hat{y}^*$.

(s) [harder] Is it reasonable to assume $f \in \mathcal{H}$? Why or why not?

Generally speaking no but there is a $h* \in H$ which is the closes possible model to 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.

(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.

The

(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?