MATH 390.4 / 650.2 Spring 2018 Homework #1t

Professor Adam Kapelner

Monday 26$^{\text{th}}$ February, 2018

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?

(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.

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

(i) [difficult] How does Silver define *out of sample*? 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. WARNING: Silver defines *out of sample* completely differently than the literature (and differently than practitioners in industry). 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.

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. In the top right quadrant, you should write "predictions" not "data" (this was my mistake in the notes). "Data / measurements" are reserved for the bottom right quadrant. The quadrants are connected with arrows. Label these arrows appropriately as well..

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

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

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

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

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

Problem 3

We are now going to investigate the aphorism "An apple a day keeps the doctor away". We will use this as springboard to ask more questions about the framework of modeling we introduced in this class.

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

(b) [easy] Is this a mathematical model? Yes / no and why.

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

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

(e) [easy] Devise a means to measure the main input. Call this x_1 going forward.

(f) [easy] Devise a means to measure the main output. Call this y going forward.

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

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

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

(j) [harder] From this point on, you only observe x_1 is in the model. 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.

(k) [harder] How did we term the functional relationship between y and x_1?

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

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

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

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

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

(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?

(t) [easy] If $f \notin \mathcal{H}$, what are the three sources of error? Write their names and provide a sentence explanation of each. Note that I made a notational mistake in the notes based on what is canonical in data science. The difference $t - g$ should be termed e as the term $\mathcal{E}$ is reserved for $t - h^*$.

(u) [harder] For each of the three source of error, provide a means of reducing the error. We discussed this in class.

(v) [easy] Regardless of your answer to what $\mathcal{Y}$ was above, we now coerce $\mathcal{Y} = \{0, 1\}$. If we use a threshold model, what would $\mathcal{H}$ be? What would the parameter(s) be?

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

Problem 4

These are questions about the linear perceptron. This problem is not related to problem 3.

(a) [easy] For the linear perceptron model and the linear support vector machine model, what is $\mathcal{H}$? Use b as the bias term.

(b) [harder] Rewrite the steps of the *perceptron learning algorithm* using b as the bias term.

(c) [easy] Illustrate the perceptron as a one-layer neural network with the Heaviside / binary step / indicator function activation function.

(d) [easy] Provide an illustration of a two-layer neural network. Be careful to indicate all pieces. If a mathematical object has a different value from another mathematical object, denote it differently.