MATH 390.4 / 650.2 Spring 2020 Homework #4

Professor Adam Kapelner

Due ....., 2020 under the door of KY604

(this document last updated 2:47pm on Wednesday 4th March, 2020)

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 Chapters 7-9 of Silver's book. 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 LaTeX. Links to instaling LaTeX and program for compiling LaTeX 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 LaTeX, print this document *including this first page* and write in your answers. I do not accept homeworks which are *not* on this printout.

NAME: _____

Problem 1

These are questions about Silver's book, chapters 7-9. For all parts in this question, answer using notation from class (i.e. $t, f, g, h^*, \delta, \epsilon, e, t, z_1, \ldots, z_t, \mathbb{D}, \mathcal{H}, \mathcal{A}, \mathcal{X}, \mathcal{Y}, X, y, n, p, x_{\cdot 1}, \ldots, x_{\cdot p}$, $x_{1\cdot}, \ldots, x_{n\cdot}$, etc.) as well as in-class concepts (e.g. simulation, validation, overfitting, etc.)

(a) [easy] Why are flu fatalities hard to predict? Which type of error is most dominant in the models?

(b) [easy] In what context does Silver define extrapolation? Give a couple examples of extraordinary prediction failures (by vey famous people who were considered heavy-hitting experts of their time) that were due to reckless extrapolations.

(c) [easy] Using the notation from class, define "self-fulfilling prophecy" and "self-canceling prediction".

(d) [easy] Is the SIR model of infectious disease under or overfit? Why?

(e) [easy] What did the famous mathematician Norbert Weiner mean by "the best model of a cat is a cat"?

(f) [easy] Not in the book but about Norbert Weiner. From Wikipedia:

> Norbert Wiener is credited as being one of the first to theorize that all intelligent behavior was the result of feedback mechanisms, that could possibly be simulated by machines and was an important early step towards the development of modern artificial intelligence.

What do we mean by "feedback mechanisms" in the context of this class?

(g) [easy] I'm not going to both asking about the bet that gave Bob Voulgaris his start. But what gives Voulgaris an edge (p239)? Frame it in terms of the concepts in this class.

Note: I will not ask questions in this assignment about Bayesian calculations and modeling (a large chunk of Chapter 8) as this is the subject of Math 341. It is obviously important in Data Science (that's why Math 341 is a required course in the new major).

(h) [easy] Why do you think a lot of science is not reproducible?

(i) [easy] Why do you think Fisher did not believe that smoking causes lung cancer?

(j) [easy] Is the world moving more in the direction of Fisher's Frequentism or Bayesianism?

(k) [easy] How did Kasparov defeat Deep Blue? Can you put this into the context of over and underfiting?

(l) [easy] Why was Fischer able to make such bold and daring moves?

(m) [easy] What metric y is Google predicting when it returns search results to you? Why did they choose this metric?

(n) [easy] What do we call Google's "theories" in this class? And what do we call "testing" of those theories?

Problem 2

These are questions...

(a) [easy] How did we define "extrapolation" in class?