Philosophy of Statistics: Bonus Homework

due on Gradescope by 11am on Tuesday June 8

Guidelines. This homework is to give people a chance to make up for missed forum posts. Each *correctly* answered question will make up for *one* missed post. It's up to you which questions, and how many, to attempt. This homework plays no role in your homework average: all it does is make up for missed forum posts. No extensions: you must submit it by the deadline for any credit.

Problem 1. In his paper about simplicity, White imagines "a certain idealized model of a regularity producing mechanism", namely, a box with a dial and pointer. He focuses on what your credences should be about the box: which function it computes and which mechanism is inside it. This question checks how well you remember White's framework. (It's similar to a problem from a previous homework, but beware: the details are slightly different.)

The square diagram below represents Sally's credences about the box.

<table>
<tr><td rowspan="8">F_1</td><td rowspan="4">F_2</td><td rowspan="2">F_2</td><td>F_1</td></tr>
<tr><td>F_2</td></tr>
<tr><td rowspan="2">F_4</td><td>F_3</td></tr>
<tr><td>F_7</td></tr>
<tr><td rowspan="4">F_3</td><td rowspan="2">F_6</td><td>F_8</td></tr>
<tr><td>F_9</td></tr>
<tr><td rowspan="2">F_7</td><td>F_{11}</td></tr>
<tr><td>F_{12}</td></tr>
<tr><td rowspan="8">F_2</td><td rowspan="4">F_4</td><td rowspan="2">F_8</td><td>F_{13}</td></tr>
<tr><td>F_{14}</td></tr>
<tr><td rowspan="2">F_9</td><td>F_{15}</td></tr>
<tr><td>F_{16}</td></tr>
<tr><td rowspan="4">F_5</td><td rowspan="2">F_{10}</td><td>F_{17}</td></tr>
<tr><td>F_{18}</td></tr>
<tr><td rowspan="2">F_{11}</td><td>F_{19}</td></tr>
<tr><td>F_{20}</td></tr>
<tr><td>M_1</td><td>M_2</td><td>M_3</td><td>M_4</td></tr>
</table>

For each function f, write down $c(f)$ and Sally's credence that the machine computes f. Does Sally satisfy the Simple Function Favoring Principle?

Sally gathers some data points, by turning the dial and observing the pointer reading. The functions consistent with her data points are $f_1, f_4, f_6, f_{12}, f_{13}, f_{19}, f_{20}$. Sally updates her credences by conditionalization, as usual. For each mechanism complexity, write down Sally's new credence that the mechanism is of that complexity. Does Sally satisfy the Simple Mechanism Favoring Principle?

Problem 2. Here is a particular crate-and-boxes problem:

First, what is Bayes decision rule (the decision rule which has the highest accuracy) in this crate-and-boxes problem?

Second, work out the overall, blue and yellow confusion tables for Bayes decision rule. (Write the entries as whole number fractions, not as decimals.)

Third, which of the four fairness properties we discussed in class, if any, does Bayes decision rule satisfy?

Problem 3. The optional reading for Class 7 was Virginia Eubanks's case study of the Allegheny Family Screening Tool (AFST) from her book "Automating Inequality". It's important, after thinking about abstract and technical questions about crate-and-boxes problems, to remember the real-life situations which motivated them. So the aim of this question is to get you to engage closely with Eubanks's case study.

Eubanks makes many different criticisms of AFST in the chapter. Describe in your own words at least eight of them in a numbered list. For each one, include a page reference and, if appropriate, a short quotation to support your point.

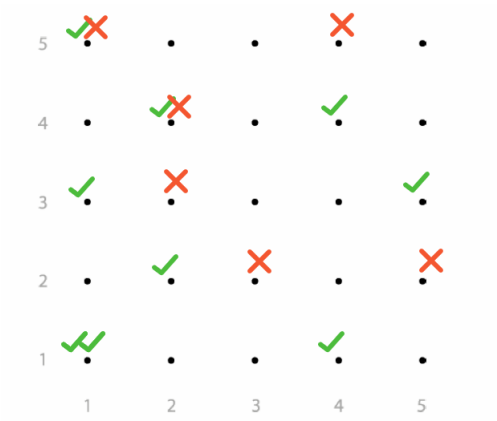
Problem 4. Imagine a trial to see whether after-school tutoring improves test performance. The trial works like this: there are 1500 students, of whom half are given tutoring and half are not; after three months, all the students take a pass/fail test.

Our data might exhibit Simpson reversals. For example, comparing the tutored group to untutored group, there might be:

- overall, a *higher* pass rate
- for students with ADHD and students without ADHD, *lower* pass rates
- for students with ADHD who live near school, students without ADHD who live near school, students with ADHD who don't live near school, students without ADHD who don't live near school, *higher* pass rates

Can you come up with such a case? Present your hypothetical data in a table, as in previous homeworks. Include the calculations of all relevant recovery rates. Remember that there are 1500 students in the study and half were given tutoring.

Problem 5. The image below represents some training data in a binary classification problem.



For each of the following learning algorithms, which decision rule does it produce given this training data?

1. Nearest Neighbor
2. 3-Nearest Neighbor
3. 1-Window

You should specify a decision rule by listing all feature vectors for which the decision rule predicts positive.