MATH 70098 Ethics in DS and AI: Part I
Checkpoint 2 Questions

Zak Varty

SOLUTIONS

This checkpoint question sheet is for you to assess your own progress through Ethics Part 1 and to identify any areas that need further clarifications.

It is recommended that you make a written attempt at all questions before solutions are made available to you.If you use code to answer any part of this assessment, this code should be clearly commented in a separate a separate `.r` or `.py` file.

This activity is expected to take approximately **90 minutes** of effort. The available marks for each question are indicated in square brackets, with a total of **26 marks** available.

---

1 Confusion at the bank [26]

A bank has developed two new methods of assessing eligibility for a business loan. The head of operations at the bank wishes to compare two new methods to their current approach. They are also concerned about the fairness of the methods, having heard anecdotal evidence that female applicants were being disadvantaged by their current approach.

You have been hired as a consultant to advise on the assessment of the accuracy and fairness of the bank's loan eligibility methods. To assist you, the bank has provided historical loan outcomes from the last year, along with the predictions made by the current and new eligibility methods. These are supplied in `loan_outcomes_NNN.csv`.

a) Construct the confusion matrix for the benchmark method and the two new methods. [3]

Solution: One mark per correct two-way table. See code output for individualised solutions.

b) Calculate the true positive rate and false positive rate for each method. Use these values to recommend which test the bank should use. You should explain the reasoning behind this recommendation. [5]

Solution:

$$\text{TPR} = \frac{TP}{TP + FN} = \frac{\#(\text{Accepted \& Repaid})}{\#(\text{Repaid})}$$

$$\text{FPR} = \frac{FP}{FP + TN} = \frac{\#(\text{Accepted \& Defaulted})}{\#(\text{Defaulted})}$$

`See code output for individualised numeric solutions`

Each pair of correct TPR and FPR (half mark for each). [1] [1] [1]
Recommending the test that is closest to (0,1) on a ROC plot (the test with the shorted 'distance' value in the R output) [1]
Justifying as having the greatest predictive power (or words to that effect) [1]

(Note: selecting a different test is okay if justified sufficiently. e.g. bank is risk-adverse and so finds a false-positive defaulting on their loan more damaging than a false-negative where no loan is given.)

c) Disaggregate the data by sex and calculate the TPR and FPR of each test by sex. In a few sentences, interpret these results for the head of operations. You should comment on whether the performance of the current test appears to differ by sex. You should also identify the test with the greatest predictive power for males and for non-males. [5]

Solution:

`See code output for individualised solutions`

Sex specific TPRs correct [2]
Sex specific FPRs correct [2]
Correct interpretation given in context of loan applications [1]

d) Write a short, contextualised description of error parity, equalised odds and equalised opportunity for the head of operations. [6]

Solution:

Two marks per metric. First mark for correct definition and second mark for each correctly put into context.

Error parity: Males and non-males should be mis-classified with the same probability/proportion. Mis-classification occurs here when someone would have repaid a loan they were not given or else failed to repay a loan that they were given.

Equalised odds: The probability of correctly accepting or rejecting the loan application should be the same, regardless of sex. This requires BOTH of the following:

$$\text{Pr(Accepted | repaid \& male)} = \text{Pr(Accepted | repaid \& not male)}$$
$$\text{Pr(Rejected | defaulted \& male)} = \text{Pr(Rejected | defaulted \& not male)}.$$

Equalised opportunity: This is a weaker version of equalised odds. The probability of correctly accepting loan application should be the same for those who will repay the loan, regardless of sex. This requires ONLY:

$$\text{Pr(Accepted | repaid \& male)} = \text{Pr(Accepted | repaid \& not male)}$$

e) Do tests 0, 1 and 2 appear to satisfy error parity by sex? Briefly describe how you might formally assess this. [3]

Solution:

Tests 0 and 1 appear to satisfy error parity since relative error rate approx 1. [1]

Test 2 has a much higher error rate for people who are not male. [1]

Formally, we should use a hypothesis test to account for sampling variability in the proportion of mis-classified individuals. [1]

f) Describe the practical and statistical issues in assessing fairness through the use of fairness metrics on this data. [4]

Solution:

Practically, it is ***not possible to satisfy all fairness metrics at once*** and so a decision has to be made as to what type of fairness the bank is concerned with. Additionally, some types of fairness ***can***

only be assessed retrospectively, because they rely on knowing the true outcome of granting the loan.

Statistically, we want to ensure that the population error rates are equal but must base this on the sample proportions. When samples are small our ***power to detect differences may be limited***. If ***sampled individuals are not representative of the population*** (e.g. only see outcomes for those who were granted loans) this must be accounted for to account for the bias it introduces.

(Note: four points given here as examples, only three valid points are required for full marks. These need not be the points listed here but should include at least one practical and one statistical issue.)

One valid statistical issue [1]
One valid practical issue [1]
One additional issue, practical or statistical [1]
Clear writing and presentation of ideas [1]