

7 Machine Learning and Real-world Data (sht25)

A new type of fraud that has been going on for the past 48 hours has been detected by a bank. The bank has manually established for each of the 120,000 transactions that occurred in this time whether it was affected by the fraud; this was the case in 14 transactions. The bank plans to deploy an automated fraud-detection system that treats the problem as binary classification.

(a) Two competing systems are to hand for the task: System A declares 6 transactions as fraud, of which 5 were indeed fraudulent. System B declares 32 transactions as fraud, out of which 12 were fraudulent. There are 5 transactions for which both systems declare fraud, out of which 4 are fraudulent.

(i) For each system, give the number of false negatives, true negatives, false positives and true positives. [2 marks]

(ii) Define and calculate the accuracy on the given data for both systems. [2 marks]

(iii) Define and calculate the precision and recall of fraud for both systems. [3 marks]

(b) The bank director decides to deploy the system with the higher accuracy, if it turns out to be significantly better than the other system. She asks you to perform a sign test to determine if this is so.

(i) Describe how you would proceed in principle. What is your null hypothesis? What are parameters N and p in the relevant formula? [Note: It is not necessary to give the formula.] [2 marks]

(ii) In how many transactions does System A beat System B and vice versa? In how many transactions do they perform identically? [4 marks]

(iii) Does accuracy, in combination with the sign test from Part (b)(i), adequately distinguish between the systems? [2 marks]

(iv) Is there a more meaningful comparison of the two systems you can offer? Can you still perform a significance test on the metric of your choice? If so, how? If not, why not? [5 marks]