CISC 7700X Final Exam

Pick the best answer that fits the question. Not all of the answers may be correct. If none of the answers fit, write your own answer.

1. (5 points) A *model* is:

   (a) A data point.

   (b) A description.

   (c) A fact.

   (d) All of the above.

2. (5 points) For last 3 years, your investment returned: {+25%, +25%, -50%}. What's the arithmetic mean of your returns:

(answer)

3. (5 points) For last 3 years, your investment returned: {+25%, +25%, -50%}. What's the geometric mean of your returns:

(answer)

4. (5 points) This technique allows assigning measures of accuracy to sample estimates of almost any statistic using random sampling methods.

   (a) Normal distribution curve with 95% accuracy

   (b) Bootstrapping

   (c) Standard deviation

   (d) 90% confidence interval

5. (5 points) A permutation test can be used for

   (a) Determining the sorting order of a randomized list.

   (b) Determining the significance

   (c) Determining the $N!$ (N-factorial) of all permutations.

   (d) Same places as Student s-Test.

6. (5 points) Both standard deviation and interquartile range measure:

   (a) The slope of the data.

   (b) The spread of the data.

   (c) The central tendency of the data.

   (d) The gradient of the data.

7. (5 points) If 1-year standard deviation is 7, then 2-year standard deviation is:

   (a) 7

   (b) 10

   (c) 50

   (d) None of the above, the answer is:

8. (5 points) You find a random widget with serial number 959569. With 50% confidence, how many widgets are out there?

   (a) somewhere between 0 and 1000000.

   (b) somewhere between 959569 and 959569*4.

   (c) at least 959569 widgets.

   (d) Not enough data to make a guess.

9. (5 points) To determine if your experiment outcome is significant, you can:

   (a) Collect a *a lot* of data.

   (b) Perform a controlled experiment.

   (c) Perform a permutation test.

   (d) Repeat the data collection and see if anything changes.

10. (5 points) If $P(x, y) \neq P(x)P(y)$ then

    (a) x is more likely than y.

    (b) x implies y.

    (c) x and y are independent.

    (d) x and y are not independent.

    (e) None of the above, answer is:

11. (5 points) If $P(y|x)P(x) \neq P(x|y)P(y)$ then

    (a) x is more likely after y.

    (b) y causes x.

    (c) x and y are independent.

    (d) x and y are not independent.

    (e) None of the above, answer is:

12. (5 points) The process of computing $P(x)$ from $P(x, y)$ is called

    (a) Bootstrapping

    (b) Generalizing

    (c) Specifizing

    (d) Marginalizing

13. (5 points) In Bayes rule: $P(x|y) = P(y|x)P(x)/P(y)$, the $P(x)$ is:

    (a) The likelihood.

    (b) The prior probability.

    (c) The posterior probability.

    (d) The posterior likelihood.

14. (5 points) In Bayes rule: $P(x|y) = P(y|x)P(x)/P(y)$, the $P(y|x)$ is:

    (a) The likelihood.

    (b) The prior probability.

    (c) The posterior probability.

    (d) The conditional probability of y given x.

15. (5 points) Conditional probability $P(y|x)$ differs from likelihood $P(y|x)$:

    (a) They're both the same.

    (b) They both sum to 1.

    (c) Probability $P(y|x)$ is a function of y, while likelihood $P(y|x)$ is a function of x.

    (d) Likelihood tells us the probability of y given x.

16. (5 points) You work for a bank credit card department. Historically, only one-tenth of one-percent (0.001) of all credit-card transactions are fraudulent. Of the fraudulent transactions, about 90% are *way-above-average-amount* for the customer, while only 0.5% of legitimate transactions are *way-above-average-amount*, about once-or-twice-a-year or so. You notice a *way-above-average-amount* transaction, what's the probability it is fraudulent?

(answer)

17. (5 points) Continuing from above: Of the fraudulent transactions, about 80% are *out-of-state*, while only 2% of legitimate transactions are *out-of-state*. You notice an *out-of-state* transaction, what's the probability it is fraudulent?

(answer)

18. (5 points) Continuing from above: You notice an *out-of-state* and *way-above-average-amount* transaction. What's the probability it is fraudulent?

(answer)

19. (5 points) Continuing from above: You notice an *out-of-state* and *way-above-average-amount* transaction. Use Naive Bayes rule to determine probability it is fraudulent:

(answer)

20. (5 points) Continuing from above: the bank refuses the transaction. The customer later confirms (via email) that it was a legitimate transaction. How would you update/alter/tweak your model to allow similar transactions in the future for that customer (assuming your model is just based on *out-of-state* and *way-above-average-amount*)?

(answer)