CISC 7700X Midterm Exam

Answers must be emailed in plain text (no formatting, no attachments). Email *must* have your *full name* at the *top*. Answers to questions must be clearly marked (question number before each answer), and be in sequence (question 1 should come before question 2, etc.).

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) The more supporting evidence we observe, the more confidence we have in the model. Suppose our model is: all fish live in water: If something is a fish, then it lives in water. Supporting evidence may consist of:

   (a) Observing a fish in water.

   (b) Observing a person on the beach.

   (c) Observing a duck flying over water.

   (d) All of the above.

2. (5 points) We make a lot of observations of A happening within 5 minutes before B. To show that A *causes* B:

   (a) We need to observe at least 1,000 instances of A happening right before B.

   (b) We need to observe at least 1,000,000 instances of A happening right before B.

   (c) Observing B without A proves that A does not cause B.

   (d) We need to conduct a controlled experiment.

3. (5 points) Smallpox: Suppose that out of 1 million people, 99% are vaccinated, and 1% are not. A vaccinated person has 1% chance of developing a reaction, which has 1% chance of being fatal. A vaccinated person has no chance of getting smallpox. An unvaccinated person has 1% chance of getting smallpox, which is fatal in 20% of the cases. Quick math shows that we can expect 99 fatalities (1000000 * 0.99 * 0.01 * 0.01) from vaccine complications, and 20 fatalities (1000000 * 0.01 * 0.01 * 0.20) from smallpox. Vaccinations kill more people than smallpox! What is wrong with the above analysis?

   (e) Answer:

4. (5 points) Fair coin flipping game: We start with \$1. Heads we win 50%, tails we lose 50%. After 2 rounds, with a fair coin, the *mean* value we will have:

   (a) \$0.25

   (b) \$0.75

   (c) \$1.00

(d) \$2.25

5. (5 points) Fair coin flipping game. We start with \$1. Heads we win 50%, tails we lose 50%. After 2 rounds, with a fair coin, the *median* value we will have:

   (a) \$0.25
   (b) \$0.75
   (c) \$1.00
   (d) \$2.25

6. (5 points) For last 3 years, your investment returned: {+35%, +35%, -70%}. Which measure of central tendency would best describe your annual return?

   (a) Arithmetic mean
   (b) Geometric mean
   (c) Median
   (d) Standard Variance

7. (5 points) If $P(x, y) \neq P(x)P(y)$ then

   (a) x is more likely than y.
   (b) x is causes y.
   (c) x and y are independent.
   (d) x and y are not independent.
   (e) None of the above, answer is:

8. (5 points) The process of computing $P(x)$ from $P(x, y)$ is called

   (a) Bootstrapping
   (b) Generalizing
   (c) Specifizing
   (d) Marginalizing

9. (5 points) If $P(x|y) \neq P(x, y)/P(y)$ then

   (a) x is more likely after y.
   (b) y is causes x.
   (c) x and y are independent.
   (d) x and y are not independent.
   (e) None of the above, answer is:

10. (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.

11. (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.

12. (5 points) If 1-year standard deviation is 5, then 2-year standard deviation is:
    (a) 5
    (b) 10
    (c) 25
    (d) None of the above, the answer is:

13. (5 points) We have two die, an 6-sided one, and an 8-sided one. We pick one at random. What's the probability we picked 6-sided die?
    (a) 1/2
    (b) 3/7
    (c) 9/25
    (d) 4/7
    (e) None of the above, the answer is:

14. (5 points) We have two die, an 6-sided one, and an 8-sided one. We pick one at random, and note the number: 4. What's the probability we picked 8-sided die?
    (a) 1/2
    (b) 3/7
    (c) 9/25
    (d) 4/7
    (e) None of the above, the answer is:

15. (5 points) We run a car dealership. From past data, 1 in 10 car loans end up bad. Our car loan application process is very detailed: we record thousands of features on each applicant, even what they are wearing, hair style, etc. After crunching past loans, we discover that bad loans were to people with mustaches 80% of the time (good loans involve a mustache only 10% of the time). A customer with a mustache walks in: use the Bayes rule to get probability that the customer will default on their loan:

(a) 1/5
(b) 4/5
(c) 5/6
(d) 1/6
(e) None of the above, the answer is:

16. (5 points) We codify the rule from previous question, and put it into production. After sometime, we discover that the default rate remains mostly unchanged. What could have gone wrong?

    (a) Obviously the implementation of the Bayes rule used wrong technology stack.
    (b) People realized we are onto mustaches and shaved them off.
    (c) If we have thousands of features to choose from, some are bound to be correlated with the result we want by pure chance.
    (d) Mustaches don't cause car loan defaults. Car loan defaults cause mustaches!

17. (5 points) Given a sample of N data points, we discover that we can fit two models, a line: $y = w_0 + w_1 x$ and a polynomial:

$$y = w_0 + w_1 x + w_2 x^2 + w_3 x^3 + w_4 x^4 + w_5 x^5$$

    The polynomial fits our training dataset 'better'. Which is true:

    (a) We'd expect the line to have lower variance, but higher bias.
    (b) We'd expect the line to have higher variance, but lower bias.
    (c) We'd expect both to have equivalent bias and variance.
    (d) We'd expect the polynomial to perform better on other samples.

18. (5 points) Given a confusion matrix, we can calculate the accuracy:

    (a) By summing all columns and rows.
    (b) By summing across the diagonal.
    (c) By removing false positives from the diagonal counts.
    (d) By comparing false negatives to false positives.
    (e) None of the above, the answer is:

19. (5 points) Given a training sample of M data points of N-dimensions: organized as a matrix $\boldsymbol{X}$ that has M rows and N columns, along with the $\boldsymbol{y}$ vector (of M numbers). We wish to fit a linear model such as:

$$y = x_0 * w_0 + x_1 * w_1 + \ldots + x_n * w_n$$

    If M is much *bigger* than N, we can solve for $\boldsymbol{w}$ via:

    (a) $\boldsymbol{w} = \boldsymbol{X}^{-1}\boldsymbol{y}$

(b) $\boldsymbol{w} = (\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{I})^{-1}\boldsymbol{y}$

(c) $\boldsymbol{w} = (\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{I})^{-1}\boldsymbol{X}^T\boldsymbol{y}$

(d) $\boldsymbol{w} = \boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{X}^T + \lambda\boldsymbol{I})^{-1}\boldsymbol{y}$

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

20. (5 points) Refer to previous question, if M is much *smaller* than N, we can solve for $\boldsymbol{w}$ via:

(a) $\boldsymbol{w} = \boldsymbol{X^{-1}}\boldsymbol{y}$

(b) $\boldsymbol{w} = (\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{I})^{-1}\boldsymbol{y}$

(c) $\boldsymbol{w} = (\boldsymbol{X}^T\boldsymbol{X} + \lambda\boldsymbol{I})^{-1}\boldsymbol{X}^T\boldsymbol{y}$

(d) $\boldsymbol{w} = \boldsymbol{X}^T(\boldsymbol{X}\boldsymbol{X}^T + \lambda\boldsymbol{I})^{-1}\boldsymbol{y}$

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