CISC 7700X Midterm 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 fact.
   (b) A description.
   (c) A data point.
   (d) All of the above.

2. (5 points) The central tendency of the data is measured by:
   (a) standard deviation
   (b) variance
   (c) median
   (d) interquartile range measure

3. (5 points) The spread of the data is measured by:
   (a) geometric mean
   (b) variance
   (c) arithmetic mean
   (d) Pearson correlation

4. (5 points) If $P(x, y) = P(x)P(y)$ then
   (a) x is more likely than y.
   (b) x is causes y.
   (c) x and y are not independent.
   (d) x and y are independent.
   (e) None of the above, answer is:

5. (5 points) The process of computing $P(x)$ from $P(x|y, z)P(y|z)P(z)$ is called
   (a) Bootstrapping
   (b) Generalizing
   (c) Specifizing
   (d) Marginalizing

6. (5 points) Suppose we have $P(A, B, C, D, E, F, G, H, I, J, K)$, where each of the $A, \ldots, K$ has values from 1 to 1000. We would like to find $P(K)$. How many loops would be required to calculate that?

   (e) Answer is:

7. (5 points) Fair coin flipping game: We start with \$1. Heads we win 50%, tails we lose 50%. After 3 rounds, with a fair coin, the *median* value we will have:

   (e) Answer is:

8. (5 points) Fair coin flipping game: We start with \$1. Heads we win 50%, tails we lose 50%. After 3 rounds, with a fair coin, the *geometric mean* value we will have:

   (e) Answer is:

9. (5 points) Fair coin flipping game: We start with \$1. Heads we win 50%, tails we lose 50%. After 3 rounds, with a fair coin, the *arithmetic mean* value we will have:

   (e) 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 prior probability.

    (b) The posterior probability.

    (c) The likelihood.

    (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) Which one of these is correct?

    (a) $P(A|B) = \frac{P(B|A)P(A)}{\sum P(A,B)}$

    (b) $P(A|B) = P(B|A)P(A)P(B)$

    (c) $P(A|B) = P(A, B)/P(B|A)$

    (d) $P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|-A)P(-A)}$

13. (5 points) Which one of these is correct?

    (a) $P(A, B, C) = P(A|B)P(B|C)P(C)$

    (b) $P(A, B, C) = P(A|B, C)P(B|C)P(C)$

    (c) $P(A, B, C) = P(A|C)P(C|B)P(B)$

    (d) $P(A, B, C) = P(A|B)P(A|C)P(B)P(C)$

14. (5 points) If $P(x|y) = P(x, y)/P(y)$ 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:

15. (5 points) On any given day, about 10% of the stocks go up by a lot, 10% go down by a lot, and 80% have no major price move. Of the stocks that go up, 60% are favortably mentioned in the press. Of the stocks that go down or stay the same, only 10% are favortably mentioned in the press. Stock XYZ is favorably mentioned in the press, use Bayes rule to find probability it will go up by a lot.

Answer is :

16. (5 points) Continuing from previous question. Of the stocks that go up, 40% have a large-insider-purchase. Of the stocks that go down or stay the same, only 5% have a large-insider-purchase. Stock XYZ has a large-insider-purchase, use Bayes rule to find probability it will go up by a lot.

Answer is :

17. (5 points) Continuing from previous question, write out the bayes rule formula to determine probability of XYZ going up after observing a favorable press coverage *and* a large-insider-purchase. Can we calculate that probability? Why?

Answer is :

18. (5 points) Continuing from previous question, write out the *naive bayes* formula to determine probability of XYZ going up after observing a favorable press coverage *and* a large-insider-purchase. Calculate the probability.

Answer is :

19. (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:

20. (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 higher variance, but lower bias.

    (b) We'd expect both to have equivalent bias and variance.

    (c) We'd expect the line to have lower variance, but higher bias.

    (d) We'd expect the polynomial to perform better on other samples.