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) Both mean and median 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.

2. (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.

3. (5 points) If $P(x, y) = P(x)P(y)$ 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:

4. (5 points) If $P(x, y) \neq P(x|y)P(y)$ 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:

5. (5 points) Which one of these is correct?
   (a) $P(A|B) = \frac{P(B|A)P(A)}{\sum P(A,B)}$
   (b) $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$
   (c) $P(A|B) = P(B|A)P(A)P(B)$
   (d) $P(A|B) = P(A, B)/P(B|A)$

6. (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.

7. (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.

8. (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.

9. (5 points) Imagine we run a bank's lending department. From past data, 1 in 5 credit applicants end up defaulting on their loans. Our loan application process is very detailed: we record thousands of data points on each applicant. After crunching past loans, we discover that bad loans were to people with mustaches 80% of the time (good loans involve a mustache only 1% 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:

10. (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 loan defaults. Loan defaults cause mustaches!

11. (5 points) We next use income brackets: from past data, 60% of loan defaults are in \$1-40k income bracket, 20% of loan defaults are in \$40-100k income bracket, 20% are in \$100-and-up bracket. 50% of good loans (non-defaulted) are from \$1-40k income bracket, 40% from \$40-100k bracket, and 10% from \$100-and-up. A customer with \$50k income walks in, according to Bayes rule what's the probability of default?

    (a) 1/9

    (a) 1/6

    (a) 1/5

    (b) Cannot be determined. Not enough information.

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

12. (5 points) Another feature that appears useful is whether the applicant has a car loan. We notice that 90% of defaulted applicants also had a car loan, while only 20% of non-default applicants had a car loan. A customer with a car loan walks in, according to Bayes rule what's the probability of default?

    (a) 82/90

    (b) 5/41

    (c) 95%

    (d) Cannot be determined. Not enough information.

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

13. (5 points) Being very clever, we first apply the income bracket model, followed by the car-loan check model. A customer with \$50k income walks in, with an existing car loan, according to Bayes rule what's the probability of default?

    (a) 5/77

    (b) 1/9

    (c) 7/55

    (d) Cannot be determined. Not enough information.

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

14. (5 points) Continuing from previous question, using Naive Bayes assumption, what's the probability of default after observing income bracket and car-loan feature?

    (a) 5/77

    (b) 1/9

    (c) 7/55

    (d) Cannot be determined. Not enough information.

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

15. (5 points) The answer to previous question is:

    (a) The exact probability of rain given the evidence.
    (b) An overestimate.
    (c) An underestimate.
    (d) Depends on whether no-car-loan implies higher income bracket.

16. (5 points) Which one of these is not a linear model? (notation tip: x^n is x raised to nth power; x_n is the nth x in the list).

    (a) $y = x_0 * w_0 + x_1 * w_1 + \ldots + x_n * w_n$
    (b) $y = x^0 * w_0 + x^1 * w_1 + x^2 * w_2 + \ldots + x^n * w_n$
    (c) $y = w_0 * e^{w_1 * x}$
    (d) $y = w_0 * x^{w_1}$
    (e) All of the above are linear.

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 higher variance, but lower bias.
    (b) We'd expect the line to have lower variance, but higher 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) Using the dataset from previous question, we wish to fit the same linear model using gradient descent. We take a guess at the initial $\boldsymbol{w}$ and start iterating: updating the $\boldsymbol{w}$ values with every element we examine. What would be an appropriate weight update rule for each $\boldsymbol{x}$?

    (a) $w_i = w_i + (y - f(\boldsymbol{x}))^2 x_i$

    (b) $w_i = w_i * \lambda(y - f(\boldsymbol{x}))x_i$

    (c) $w_i = w_i - \lambda(y - \boldsymbol{x}^T\boldsymbol{w})x_i$

    (d) $w_i = w_i + \lambda(y - \boldsymbol{x}^T\boldsymbol{w})x_i$

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