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) Data Science is:
   (a) Deduction of true facts using logic and math.
   (b) Describing data using statistics.
   (c) Using inference to induce models from data.
   (d) Using Python, Hadoop, and Spark to work with data.

2. (5 points) A *model* is:
   (a) A fact.
   (b) A data point.
   (c) A description.
   (d) All of the above.

3. (5 points) Both mean and median measure:
   (a) The spread of the data.
   (b) The central tendency of the data.
   (c) The slope of the data.
   (d) The gradient of the data.

4. (5 points) Both standard deviation and interquartile range measure:
   (a) The spread of the data.
   (b) The central tendency of the data.
   (c) The slope of the data.
   (d) The gradient of the data.

5. (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 independent.
   (d) x and y are not independent.
   (e) None of the above, answer is:

6. (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:

7. (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:

8. (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) $P(A|B) = \frac{P(B|A)P(A)}{P(B)}$

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

9. (5 points) Which one of these is correct?

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

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

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

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

10. (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.

11. (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.

12. (5 points) From our past experience, we know it rains 1 in 5 days. When it rains, we observe 90% of the people carry umbrellas. When it's not raining, only 10% of the people carry an umbrella. We're in a basement (no windows), and we observe someone walking in with an umbrella. Using Bayes rule, what's the probability that it's raining?

    (a) 0.95

(b) 9/13

(c) 4/13

(d) 10/31

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

13. (5 points) Continuing from previous question, we next observe someone wearing a rain-jacket. When it rains, we know about 90% of the people wear rain-jackets, and only about 10% of the people wear rain-jackets when it's not raining. Use the Bayes rule to find probability of rain given this additional evidence.

    (a) 81/85

    (b) 9/13

    (c) 4/13

    (d) 10/31

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

14. (5 points) Continuing from previous question, using Naive Bayes assumption, what's the probability of rain now that we've observed both umbrella *and* a rain-jacket?

    (a) 81/85

    (b) 9/13

    (c) 4/13

    (d) 10/31

    (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) Would change if we first observed rain-jacket followed by umbrella.

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