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.

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.).

Email must arrive by midnight on 2020-12-17.

1. (5 points) A *model* is:
   (a) A data point.
   (b) A fact.
   (c) A description.
   (d) All of the above.

2. (5 points) The more supporting evidence we observe, the more confidence we have in the model. Suppose our model is: *all cakes are sweet*: If something is a cake, then it is sweet. Supporting evidence may consist of:
   (a) Observing a sweet cake.
   (b) Observing a sour apple.
   (c) Observing a sweet apple.
   (d) All of the above.

3. (5 points) You find a random widget with serial number 54321. With 50% confidence, how many widgets are out there?
   (a) somewhere between 0 and 100000.
   (b) somewhere between 54321 and 54321*4.
   (c) at least 1000000 widgets.
   (d) Not enough data to make a guess.

4. (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.

5. (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.

6. (5 points) There is about 0.05% chance of having a certain disease19. About 80% of the people with the disease experience symptoms of high fever and cough. About 5% of those that have symptoms end up testing positive. Joe wakes up one morning with a high fever and cough. What is the probability that Joe will test positive for disease19?

   (e) answer is:

7. (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.

8. (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:

9. (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:

10. (5 points) We run a bank's lending department. From past data, 15% of loans end up bad. 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 60% 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:

(e) answer is:

11. (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!

12. (5 points) We next use income brackets: from past data, 70% of loan defaults are in $1-40k income bracket, 20% of loan defaults are in $40-100k income bracket, 10% are in $100-and-up bracket. 60% of good loans (non-defaulted) are from $1-40k income bracket, 35% from $40-100k bracket, and 5% from $100-and-up. A customer with $50k income walks in, according to Bayes rule what's the probability of default?

   (e) answer is:

13. (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 25% 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?

   (e) answer is:

14. (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?

   (e) answer is:

15. (5 points) Continuing from previous question, using Naive Bayes assumption, what's the probability of default after observing income bracket and car-loan feature?

   (e) answer is:

16. (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

17. (5 points) The process of computing $P(x)$ from $P(x, y)$ is called

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

(d) Marginalizing

18. (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:

19. (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:

20. (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: