CISC 7700X Midterm Exam

- Answers must be emailed in plain text (no formatting, no attachments).
- Email must arrive arive before midnight on 2021-10-27.
- Email *must* have your *full name* at the *top*.
- Email subject must be "CISC 7700X Midterm Exam".
- Email to: `alex@theparticle.com`.
- 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. Each question is worth 5 points. If none of the answers fit, write your own answer.

1. In Bayes rule: $P(x|y) = P(y|x)P(x)/P(y)$, the $P(x|y)$ is:
   (a) The likelihood.
   (b) The prior probability.
   (c) The posterior probability.
   (d) The conditional probability of y given x.
2. In Bayes rule: $P(x|y) = P(y|x)P(x)/P(y)$, the $P(y|x)$:
   (a) Is the prior probability.
   (b) Is a conditional probability.
   (c) Sums to 1.
   (d) Can often be estimated from past observations.
3. In Bayes rule: $P(x|y) = P(y|x)P(x)/P(y)$, the $P(x)$:
   (a) Is the prior probability.
   (b) Is a conditional probability.
   (c) Is the posterior probability.
   (d) The likelihood.
4. What measures spread of data?
   (a) Geometric mean
   (b) Arithmetic mean
   (c) Variance.
   (d) Median
5. What characterizes a random process vs chaotic process.

(a) Random is predictable short term, chaotic is predictable long term.

(b) Random is predictable long term, chaotic is predictable short term.

(c) Both are completely unpredictable.

(d) Combining random and chaotic processes creates a process that is predictable short therm and long term.

6. Medians vs Means:

   (a) Medians are more robust to noise.

   (b) Medians are easier to calculate than means.

   (c) Both measure the spread of data.

   (d) Both means and medians are the same.

7. Central Limit Theorem essentially says:

   (a) The result of any experiment is normally distributed.

   (b) The mean values are normally distributed.

   (c) All probability distributions are normal.

   (d) Variance is normally distributed.

8. The k-NN model:

   (a) Fits a hyperplane to the k nearest training instances.

   (b) Is a k level neural network.

   (c) Uses k nearest training instances to vote/predict the class label.

   (d) Builds a decision tree k levels deep.

9. Geometric mean (as opposed to an arithmetic mean) is:

   (a) Is a measure of error around variance.

   (b) Is a geometric shape around the arithmetic mean.

   (c) Is almost the same as arithmetic mean.

   (d) Useful in situations that involve compounding.

10. For a matrix X, with N rows and M columns, what's the size of $X^T X$?

    (e) Answer:

11. For a matrix X, with N rows and M columns, what's the size of $X X^T$?

    (e) Answer:

12. For a matrix X, that has many more columns than rows, which one is easier to invert: $X^T X$ or $X X^T$?

    (e) Answer:

13. When comparing two business entities (e.g. companies, customers, suppliers) what is the best distance metric to use?

    (a) Euclidean Distance
    (b) Manhattan Distance
    (c) Chebyshev Distance
    (d) Mahalanobis Distance
    (e) Answer:

14. A high correlation between A and B tells us:

    (a) A causes B
    (b) B causes A
    (c) C causes both A and B
    (d) that A and B happen together, perhaps by coincidence.
    (e) Answer:

15. One reason to use ratios of features instead of actual features is that:

    (a) There is less chance of coincidence in training.
    (b) Ratios are generally invariant to scale.
    (c) Ratios are normally distributed, according to central limit theorem.
    (d) Ratios have higher variance and lower bias.
    (e) Answer:

16. We work at a bank, and have access to: credit card application AND resulting credit behavior a year later. We use this labeled data to build a model to approve credit card applications. It works great on our training data, but fails very badly in real life. What went wrong?

    (a) Past data may not predict the future.
    (b) The new model is trained on approved credit card applications.
    (c) Times have changed, and new model may not be appropriate in today's world.
    (d) Too many parameters made the new model seem better than it was.

17. We are fans of Marvel movies. Whenever a new movie comes out, there's a 70% chance we'd love it (from past data). Of the Marvel movies we love, 40% of them have IronMan. Of the movies we don't like, only 10% of them have IronMan[1]. A new movie with IronMan is coming out, use Bayes rule to estimate probability we'd love it.

    (e) Answer:

[1] Ironman II?

18. We also love Spiderman movies. Of the Marvel movies we love, 20% of them have Spiderman. Of the movies we don't like, only 10% of them have Spiderman. A new movie with Spiderman is coming out, use Bayes rule to estimate probability we'd love it.

    (e) Answer:

19. A new movie with both IronMan and Spiderman is coming out. Use Bayes rule to estimate probability we'd love it.

    (e) Answer:

20. A new movie with both IronMan and Spiderman is coming out. Use Naive Bayes to estimate probability we'd love it.

    (e) Answer: