Self-Assessment Sheet

Management & Data Science

Leuphana University of Lüneburg

Dear applicant,

The Data Science Master Program at Leuphana University of Lüneburg is of interest to a very heterogeneous crowd and we are enjoying strong demand. Accepted students need to share a common theoretical ground to make the program a success. This sheet contains basic exercises and questions that you may use to test yourself. The tasks cover topics in maths, statistics, and computer science that we expect students to bring along to ensure a common ground. You'll probably find some of the tasks easy, some may ring a bell, and some others leave you clueless. Catch up the blanks to be well prepared before you join us in Lüneburg!

Best,
Ulf Brefeld

Professor for Machine Learning
Leuphana University of Lüneburg
brefeld@leuphana.de

Counting

1. Consider the letters A-F. How many different arrangements are there for which:
   - A and B are next to each other?
   - A is before B?
   - A is before B and B is before C?
   - A is before B and C is before D?
   - A and B are next to each other and C and D are also next to each other?
   - E is not last in the line?
2. How many 7-digit phone numbers are possible, assuming that the first digit cannot be a 0 or a 1?

Probability

3. If you toss a fair coin five times what is the probability of observing five heads?
4. Consider a fair dice. What is the probability of a six given that you already observed a six in the roll before?
5. Consider a fair dice. What is the probability of observing a six three times in a row?
6. Let A and B be events. Interprete the following equations:
   - $\mathbf{P}(A, B)$
   - $\mathbf{P}(A \cup B)$
   - $\mathbf{P}(A \cap B)$
   - $\mathbf{P}(A|B)$
7. What is Bayes' Theorem?
8. Assume that the following holds true:
   - 1% of people have a certain genetic defect
   - 90% of the tests detect the defect (true positives)
   - 9.6% of the tests are false positives

   What is the probability of a person getting a positive result, given they have the defect?
9. Assume that the weight of a one-year-old is normally distributed with a mean of 9.5 kg and a standard deviation 1.1 kg. Without using a calculator, state whether the following conclusions are right or wrong:

- Not more than 20% of the girls should have a weight less than 8.4 kg
- 95% of the girls have a weight between 7.3 kg and 11.7 kg
- 3% of the girls have a weight more than 12.8 kg

Logarithm

10. Compute the following logarithms without using a calculator:
    - $\log_2(16)$
    - $\log_5(25)$
11. Are the following statements true or false?
    - $\log_6(5y) = \log_6(5) + \log_6(y)$
    - $\log_3(x) - \log_3(y) = \log_3(\frac{x}{y})$
    - $4\log_5(2) = \log_5(16)$

Analysis

12. Compute the first derivatives of the functions stated below:
    - $f(\lambda) = 20\lambda^5$
    - $g(z) = 2z^3 + 4z^2$
    - $h(\phi) = 7\phi + 10$
    - $u(\theta) = 4\theta^5 - 2\sqrt{\theta} + \frac{3}{\theta} + 8$
    - $v(\gamma) = \ln(2 + 3\gamma^2)$

Linear Algebra

13. Compute the Eucledian distance of the vectors $(3, -2, 5)$ and $(2, 2, 3)$.
14. What is a unit vector?
15. What is the unit vector of $(2, 5, 1)$?
16. Let $\mathbf{x} = (2, 5, 1)$ and $\mathbf{y} = (-1, 3, 2)$ be two vectors in $\mathbb{R}^3$.
    - Compute $\mathbf{x}^\top \mathbf{y}$.
    - Compute $\mathbf{x}\mathbf{y}^\top$.
    - Are $\mathbf{x}$ and $\mathbf{y}$ orthogonal?
17. Let $A = \begin{pmatrix} -1 & 3 \\ 2 & -6 \end{pmatrix}$ be a matrix.

- What is the rank of A?
- What is the determinant of A?
- What is the trace of A?
- Are the columns of A linearly independent?

18. Rewrite the following system of equations in matrix notation and solve it accordingly.

$$2x + 5y = 10$$
$$3x + 4y = 24$$

19. What is an orthogonal matrix?

Miscellaneous

20. What is a Venn diagram?
21. Compute the values of the following terms:
    - $\binom{10}{6}$
    - $\frac{9!}{2!7!}$
    - $\sum_{k=1}^{5} k^2$

Statistics

22. What is the difference between mean, median and modus?
23. What is the difference between variance and standard deviation?
24. What is nominal, ordinal and interval data? Give one example each.
25. What is an outlier?
26. What is the difference between a histogram and a kernel density estimate?
27. What is a box plot? Describe its components.

Computer Science

28. What is the difference between a linear search and binary search? What are the time complexities?
29. When would you use a `for`-loop and when a `while`-loop?
30. Explain the concept of recursion.

31. What is pseudo code good for? (Can you read and write pseudo code?)
32. Implement the Fibonacci series in your favorite programming language.