

# Exam Machine Learning & Data Science II Practice

Neighbours

Lecturer: Daniel T. McGuiness, Ph.D

SEMESTER: WS 2024

DATE: 16.12.2024

TIME: 01:00-02:30 pm

First and Last Name

Student Registration Number

|                |                             |   |
|----------------|-----------------------------|---|
| Grading Scheme | $\geq 90\%$                 | 1 |
|                | $\leq 80\%$ and $\geq 90\%$ | 2 |
|                | $\leq 70\%$ and $\geq 80\%$ | 3 |
|                | $\leq 60\%$ and $\geq 70\%$ | 4 |
|                | $\leq 60\%$                 | 5 |

Result:

\_\_\_/ max. \_\_\_ points

Grade:

Student Cohort BA-MECH-22

Study Programme Bachelor - Mechatronik, Design & Innovation

Permitted Tools Practice these questions well before the exam.

## Important Notes

### Unnecessary Items

Place all items not relevant to the test (including mobile phones, smartwatches, etc.) out of your reach.

### Identification (ID)

Lay your student ID or an official ID visibly on the table in front of you.

### Examination Sheets

Use only the provided examination sheets and label each sheet with your name and your student registration number. The sheets be labelled on the front. Do not tear up the examination sheets.

### Writing materials

Do not use a pencil or red pen and write legibly.

Good Luck!

| Exam Question      | Maximum Point | Result |
|--------------------|---------------|--------|
| Practice Questions | 100           |        |

**[Q1] Practice Questions** 100

1. Please name four (4) different types of applications where the use of machine learning is could be useful. Please describe the examples with sufficient detail with relevant tools to tackle each example you have provided.
2. What is a labelled training set? Please give 2 relevant examples.
3. What are the two (2) most common supervised tasks? Please give industrially relevant examples.
4. Please name (4) common unsupervised tasks? Please give examples with sufficient detail.
5. What type of algorithm would you use to segment your customers into multiple groups ? Please justify your reasoning.
6. Would you frame the problem of spam detection as a supervised learning problem of an unsupervised learning problem? Please justify your answer.
7. What is an online learning systems ? What are its advantages and disadvantages compared to other learning systems ?
8. What is precision and recall ? Is it possible to have a system which has both perfect precision and recall? Please give an example of a perfect recall and a perfect precision.
9. Which linear regression training algorithm would be the best choice if you have a training set with millions of features?
10. Can gradient descent get stuck in a *local minimum* when training a logistic regression model? Please explain it with sufficient detail.
11. Do all gradient descent algorithms lead to the same model, provided you let them run long enough? If so, why ?
12. Suppose you use batch gradient descent and you plot the validation error at every epoch. If you notice that the validation error consistently goes up, what is likely going on? What would be the cause and how can this be fixed?
13. What is the fundamental idea behind support vector machines and what is a support vector? Please explain both these questions with sufficient detail and draw diagrams or write equations if necessary.
14. Why is it important to scale the inputs when using support vector machines ? What down-sides would occur if no scaling is done?

15. What is the point of using the kernel trick? Please give a brief explanation of its working principle.
  16. Please explain the concept of Gini impurity. Is a node's Gini impurity generally lower or higher than its parent's ? Please explain both these questions with sufficient detail and write down equations if necessary.
  17. If a decision tree is over-fitting the training set, is it a good idea to try decreasing `max_depth`?
  18. If a decision tree is under-fitting the training set, is it a good idea to try scaling `max_depth`?
  19. If you have trained five (5) different models on the exact same training data, and they all achieve 95% precision, is there any chance you can combine these models to get better results? If so, how? If not, why?
  20. What is the difference between **hard** and **soft** voting classifiers? Please give examples in your explanation.
  21. What is the benefit of out-of-bag classification?
  22. If the gradient boosting ensemble over-fits the training set, should you increase or decrease the dataset?
  23. Please explain the concept of curse of dimensionality and use examples in your explanation.
  24. When a dataset dimension has been reduced from  $n$  to  $n - 1$  is it possible to go back (i.e., has information been lost) ? If so, why? If not why?
  25. Please explain the concept of clustering with two (2) algorithms and examples.
  26. What are the main applications of clustering algorithms? Please give examples of where clustering algorithms would be the preferred options over other machine learning algorithms.
  27. Please give a scenario where active learning would be useful ? If there is an application please give an example on how you would implement it into your workflow.
-