

Exam ML & DS II Practice

Neighbours

Lecturer: Daniel T. McGuiness, Ph.D

SEMESTER: WS 2025

DATE: 12.12.2025

TIME: 09:00-10:30 pm



First and Last Name

Student Registration Number

Grading Scheme	$\geq 90\%$	1
	$\leq 80\% \text{ and } \geq 90\%$	2
	$\leq 70\% \text{ and } \geq 80\%$	3
	$\leq 60\% \text{ and } \geq 70\%$	4
	$\leq 60\%$	5

Result:

____/ max. 100 points

Grade:

Student Cohort BA-MECH-23

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!

Question	Maximum Point	Received Point
Practice Questions	100	
Sum	100	

[Q1] Practice Questions 100

1. Please name four (4) different types of applications where the use of machine learning could prove itself to 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 where labelled training set could be used and briefly discuss its advantages/disadvantages with comparison to unlabelled training sets.
3. What are the two (2) most common supervised learning methods? Please give industrially relevant examples and explain the chosen methods in sufficient detail.
4. Please name common unsupervised tasks? Please give examples with sufficient detail.
5. What is normalisation and why is it important when pre-processing data for use in ML training?
6. You are given a set of data regarding login to a news website and are tasked with determining age/gender/occupation of the users entering the website. The only information you have are of their **IP address** and the time in which they logged in to their website. Based on this information, please explain the workflow in which this problem could be tackled (i.e., Data Cleaning, type of ML method)
7. You are given a set of data on weather reports from 2015–2025 and are tasked with predicting future weather pattern within the region. Based on this information please explain the pipeline you would use to apply a ML algorithm.
8. Working as a bank employee you are tasked with detecting credit card fraud. As an engineer with a background in ML you can automate this task. Please choose a method for tackling this problem, explain why that method would work for this task.
9. What type of algorithm would you use to segment your customers into multiple groups? Please justify your reasoning.
10. As a hobbyist you like to climb mountains and collect local flowers, after you have amassed a certain amount of them you would like to classify them. However, you are not knowledgeable enough to differentiate them using traditional methods and therefore think of using ML. Please explain what kind of learning method would fit into this problem and then choose an algorithm to tackle it along with explanation of the algorithm.

11. You work in the post office and tasked with detecting handwritten numbers on sent letters. You suggest to use ML to automate the process. After discussing this with colleagues you are recommended to use a Single Layer Perception with Regression to detect digits on a paper. Is this a valid approach? If yes, please elaborate. If not, then please suggest an alternative method for a better approach
12. Would you consider the problem of spam detection as a supervised learning problem or an unsupervised learning problem? Please justify your answer.
13. What is an online learning systems? What are its advantages and disadvantages compared to other learning systems?
14. 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.
15. What options would you consider if your Support Vector model keeps overfitting?
16. Why is it important to scale the inputs when using support vector machines? What downsides would occur if no scaling is done?
17. What is the point of using the kernel trick? Please give a brief explanation of its working principle.
18. 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. Finally please give an alternative to Gini impurity which can be used instead.
19. If a decision tree is over-fitting the training set, is it a good idea to try decreasing `max_depth`?
20. If a decision tree is under-fitting the training set, is it a good idea to try scaling `max_depth`?
21. 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?
22. Please explain briefly the concept of law of large numbers and its relation with machine learning algorithms.
23. What is the difference between **hard** and **soft** voting classifiers? Please give examples in your explanation.
24. What is random forest classification? Briefly explain its operational principle.
25. What is the benefit of out-of-bag classification?
26. What is feature importance and why is it important? Which algorithm could be used to generate a feature importance map of a dataset?

27. What is AdaBoost classifier? Briefly explain its operational principle.
28. If the gradient boosting ensemble over-fits the training set, should you increase or decrease the dataset?
29. Please explain the concept of curse of dimensionality and use examples in your explanation.
30. 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?
31. What is Principle Component Analysis? Briefly explain its operational principle and its relevancy in machine learning.
32. Please explain the concept of clustering with two (2) algorithms and examples.
33. 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.
34. 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.
35. What makes Neural Networks different compared to other machine learning algorithms such as Support Vector machines, Random forest, or Decision Trees?
36. What is a perceptron? Please draw a diagram and explain its parts and write all relevant equations of it.
37. Please explain line by line the following code.

```
1 model = tf.keras.Sequential([
2     tf.keras.layers.Flatten(input_shape=[28, 28]),
3     tf.keras.layers.Dense(300, activation="relu"),
4     tf.keras.layers.Dense(100, activation="relu"),
5     tf.keras.layers.Dense(10, activation="softmax")
6 ])
```

Sample
pyton

38. What is an activation function? why is it needed?
39. Please explain three popular activation functions, write their function and briefly explain their advantages, disadvantages when used in a network.
40. Please explain the XOR problem and why is it not possible to solve it using a single Perceptron?
41. Why is it required for all weights of a neural networks to be randomised? What would occur if they were all set the same during initialisation. In this context does biases also need to be treated the same way or is it OK if they are kept the same value at the beginning of training?
42. Can a neural network produce non linear output if the activation function is a linear one? If not please explain and show mathematically.

43. Consider a CNN composed of three convolutional layers, each with 3×3 kernels, a stride of 2, and **same** padding. The lowest layer outputs 100 feature maps, the middle one outputs 200, and the top one outputs 400. The input images are RGB images of 200×300 pixels:
- What is the total number of parameters in the CNN?
 - If we are using 32-bit floats, at least how much RAM will this network require when making a prediction for a single instance?
 - What about when training on a mini-batch of 50 images?
44. What is a filter in the context of CNNs? Why are they used?
45. Please explain in significant detail the concept of a feature map.
46. Please explain the concept of pooling and explain its relevance in terms of CNNs.
47. An image tensor has the following information:

`TensorShape(5, 680, 480, 12)`

Please explain what these numbers mean and give a possible source of the image.

48. An image of size (600×800) is applied the following convolution layer:

```
conv_layer = tf.keras.layers.Conv2D(filters=32, kernel_size=5, padding="same")
```

Please calculate the resulting processed image size in terms of its dimensions.

49. The following architecture is given:

```
1 ...
2 model = tf.keras.Sequential([
3     DefaultConv2D(filters=64, kernel_size=7, input_shape=[28, 28, 1]),
4     tf.keras.layers.MaxPool2D(),
5     DefaultConv2D(filters=128),
6     DefaultConv2D(filters=128),
7     tf.keras.layers.MaxPool2D(),
8     DefaultConv2D(filters=256),
9     DefaultConv2D(filters=256),
10    tf.keras.layers.MaxPool2D(),
11    tf.keras.layers.Flatten(),
12    tf.keras.layers.Dense(units=128, activation="relu",
13                          kernel_initializer="he_normal"),
14    tf.keras.layers.Dropout(0.5),
15    tf.keras.layers.Dense(units=64, activation="relu",
16                          kernel_initializer="he_normal"),
17    tf.keras.layers.Dropout(0.5),
18    tf.keras.layers.Dense(units=10, activation="softmax")]
19 ...
```

Sample
python

Based on this code snippet, please explain:

- What is the image size used in training the model,
- Why is the each successive filter size increases with each successive layer?
- Why do we need to **flatten** the output to feed it into a neural network?
- What does **softmax** activation function do?
- What does **MaxPool2D()** do in this context?

- f) Please explain the reasoning behind the use of `tf.keras.layers.Dropout(0.5)`.
50. Please explain the architecture of AlexNet?
51. What is Data Augmentation? Why is it generally used in training networks? Why is it not recommended to use white noise in the training process?
52. Please explain the LeNet architecture.
53. What is the concept of transfer learning? Briefly explain.
54. What is the difference between padding `same` and `valid`?
55. Please explain the concept of dropout in the context of CNNs and explain its usage.
56. Please write down three (3) common CNN architectures and briefly explain them.
57. How would you handle the problem of overfitting in the context of CNNs?
58. What is the concept of “local response normalization” (LRN), and is it still commonly used in CNNs?
59. What is a Local Normalisation Layer? Why is it used?
60. Why was AlexNet revolutionary in comparison to other Machine Vision algorithms at the time?