



Course DIT 821 Software Engineering for AI Systems

Summer Course 2020
Exam preparation (II)



Exam Overview

- A combination of a written examination run remotely using Zoom.
 - Each student will get questions and will have 60 minutes to write an programming assignment
 - The students will have different assignments
 - You will be able to use material from web
 - You are not allowed to have any communication for producing the results.
 - You will submit the answer within a given period.
- After the submission you will have an oral discussion with the teachers
 - 15 minutes about the assignment (Pier and Lucy)
 - 15 minutes for two theoretical questions (Ivica)
- The correctness of the answer will give the grade (VG, G, or failed)



Exam Process I

- When login you will be placed in a (virtual) room
- Individual channel/rooms of Zoom will be used
 - Video ON
 - Showing your ID (we will make a screenshot)
 - You can use a virtual background, or a background you select
 - The teachers will randomly login to the channels
 - You can send questions via room chat
- Written part of the examination (60 minutes)
 - You will get one simple programming problem to solve
 - Means to write answers: Jupyter Notebook
- Oral part of the examination
 - After submission will wait ca 15 minutes for the oral exam:
 - Part 1 – discussion of the assignment (15 min)
 - Part 2 – theoretical questions (15 min).



Schedule – Alternative – 2 hours for the written part

Student	written exam	oral - the program as.	oral exam theory
		Pier & Lucy	Ivica
Jude Gyimah	12:30-13:30	13:45-14:00	14:00-14:15
Federica Comuni	12:45-13:45	14:00-14:15	14:15-14:30
Raghwendra Singh	13:00-14:00	14:15-14:30	14:30-14:45
Filip Norberg Åslin	13:15-14:15	14:30-14:45	14:45-15:00
Osman Osman	13:30-14:30	14:45-15:00	15:00-15:15
Shab Pompeiano	13:45-14:45	15:00-15:15	15:15-15:30
	14:00-15:00	15:15-15:30	15:30-15:45
Alexander Andersson	14:15-15:15	15:30-15:45	15:45-16:00
Erik Tingström	14:30-15:30	15:45-16:00	16:00-16:15
Haider Ali	14:45-15:45	16:00-16:15	16:15-16:30
Himanshu Chuphal	15:00-16:00	16:15-16:30	16:30-16:45
Ahmad Abdulal	15:15-16:15	16:30-16:45	16:45-17:00
Ahmed Groshar	15:30-16:30	16:45-17:00	17:00-17:15
Nawar Aghi	15:45-16:45	17:00-17:15	17:15-17:30
Nicholas Fisher	16:00-17:00	17:15-17:30	17:30-17:45

Note:

- The first participant starts 12:30
- You should login into zoom some 5 minutes before the start
- Setup your video and prepare ID
- We will use a screenshot to keep the Identification.
- At the given schedule the assignment from CANVAS will be available
- If you wish we can record the oral part of the exam



Examination criteria

- U – if **two** assignment are not answered or the student does not show understanding of the submitted assignment
- G – if a student shows understating and knowledge of at least two assignments, but the presentation is not complete, or the student cannot answer to some additional questions.
- VG - if a student shows a detailed knowledge of all three assignments, and is able to answer to discussion questions like:
 - what would happen if...
 - What would be alternative solutions
- The grade will be document (in Canvas) and then reported to Ladok



Example of questions – theoretical questions

Explain polynomial regression, hypothesis, cost function	Explain classification principles when using neural networks
Explain underfitting and overfitting	explain principles of convolution operation
Explain how the regularization factor influence the overfitting	Explain principles of reinforcement learning
Explain a process of modelling, validation and testing, and cost functions when changing size of training data set, regularization factor, degree of polynomial	Explain main principles of reinforcement learning
Explain confusion matrix principles	Explain Markov Decision Process
Explain meaning of accuracy, recall, precision and possible actions to increase the quality of the prediction	Explain the principles of the Bellman Equations
Explain principles for classification, hypothesis and its interpretation	Unig the example "Career Options" expalin terms Reward, Return, Value of state, value function
Explain logistic regression, and decision boundary	Describe the different stages of machine learning workflow.
Explain logistic regression cost function	You are a team lead responsible for developing a machine learning system that helps bank managers understand which loans to approve. Your team is provided with a data set containing over 3000 records of loan applicants. This dataset contains decision information of whether the loan was approved or not, as well as others (e.g., age, profession and credit) for each applicant. Describe the how you would go about developing such a system.
Explain how to calculate minimum of logistic regression function	Describe what feature "usefulness" means?
Explain how to deal with underfitting and overfitting for logistic regression function	Give one example when it is useful to apply Log Transform in dataset?
Explain multi-class classification	What is the difference between filters, wrappers and embedded feature selection methods?
Explain decision tree principles and decision boundary	In data labelling, what is inter-annotator agreement and how would you measure it?
Explain the principles of building a decision tree	What is data drift? Give an example and state how it can be detected in a deployed ML-enabled application?
Explain the principles for splitting attributes	Which are the basic principles of AI ethics - explain them
Explain the principles of random forest	
Explain the k-mean algorithm	
Explain principles of artificial neural network (perceptron, activation function, the network)	



Example of questions – providing solutions

You are expected to be able to run jupyter notebook and you will be given a code snippet that you would need to write or modify to find a solution

Or you can use of other means of tools to get some calculations (calc, excel/google sheet)

For given dataset find the hypothesis and calculate the cost function

Find the best value of the regularization parameter for a given data training and test set

For a given dataset find which selection of features gives the best result

Giving the accuracy of an instrument, and a distribution of positive/negative examples, calculate the probability of true positive for positive-tested (example alcohol test).

For given dataset find clusters if 3 clusters are supposed to be found

For given dataset estimate how many clusters you would define to achieve optimal results (use elbow principle)

Suppose you want to optimize regularization parameter. For given dataset find the optimal value of the parameter.

For a give dataset implement the algorithm one vs. all

Analyse dataset, and prepare dataset (clean “dirty data – eliminate outliers, manage missing values...”) for training

For a given data set provide training and its analysis



Example of questions

1. Solve the problem in specified Jupyter Notebook (see zip file)
2. Explain underfitting and overfitting
 - Present examples of underfitting and overfitting
 - When underfitting and overfitting typically appear
 - What are the means for avoid them
 - how can you use the regularization parameter for to find the best value?
3. Describe the ML workflow
 - Present the workflow
 - Describe short the actives
 - Suppose that you know that data in operation will be different from training data. Which measures you will use to avoid wrong results in the operation?