

Applied Optimization and Machine Learning – 2022/2023

Retake Exam – 11/07/2023

Full name: _____

Student Number: _____

This exam is divided in two parts. Part 1 is a set of short, independent questions on topics in simulation, optimization and machine learning. Part 2 contains three small cases which will require longer answers. Part 1 is worth 40% of the grade. Part 2 is worth 60% of the grade.

This is a long exam. If you want to answer all questions, you will need to advance fast. Be economical about your time. I advice you to not get stuck in one question if you are not sure about the answer. Locate the ones that are easy and feel obvious to you and focus on them first.

The space you are given to provide answers to each question has been designed intentionally. If a question only has a small area to answer, that is because a brief, to the point answer is expected. Don't try to find space somewhere else to extend the length of your answer.

Your exam includes a few pages of blank, draft paper at the end. You can use it for anything you need. I won't review the contents of it.

Please, use a pen. I will ignore anything written with a pencil.

I wish you good luck.

Part 1

1. How can we model historical data to use it in our simulations?
2. Provide two advantages of simulation modeling over linear/integer programming.
3. If one wants to understand the behaviour of a system by simulating it, is running the simulation once typically enough to extract conclusions from it? Why?
4. What are the usual advantages of commercial simulation packages over open-source ones?

5. What is simulation-based optimization?

6. We discussed how random search is typically a poor algorithm choice to run simulation based optimization. But still, it is very useful for one thing. What is it?

7. What are heuristics and meta heuristics? How are they different?

8. Imagine we run a genetic algorithm without mutation. What will happen? Similarly, explain what would happen if we selected individuals randomly instead of by their fitness.

9. How can a linear programming model be solved?

10. What are binary variables useful for?

11. Why do we generally prefer to simplify problems into linear or integer programs?

12. What is a solver?

13. Describe three criteria you would use to pick a solver for solving a specific optimization problem.

14. What is sensitivity analysis?

15. Provide two circumstances where linear or integer programming would be a better tool than simulation-based optimization.

16. Explain the difference between Supervised and Unsupervised Machine Learning.

17. Why do we say that Supervised Machine Learning is an optimization problem in disguise?

18. In classification problems, what is an unbalanced dataset? What challenges does it pose?

19. What is k-fold crossvalidation? Why is it useful?

20. What is information gain in the context of decision tree building?

21. How can we provide a numerical estimate of the certainty of a prediction made by a Random Forest model?

22. Compare precision and recall. When is each of them the right metric?

23. How can you know if a Machine Learning model is underfitting?

24. What would you do if your Machine Learning model is overfitting?

25. Compare bagging and boosting, providing the advantages and weaknesses of each.

Part 2

Simulation

You are working for a taxi fleet company. The operations manager of the fleet is responsible for everything on day to day operations of the service provided to customers. He needs to take decisions such as how many vehicles should be active every day and hour and where should they be positioned. He is also in charge of dealing with emergencies and unexpected situations, such as vehicles having accidents or areas of the city being shutdown to cars.

Propose metrics that are relevant to measure the success of the fleet operation

Are there trade-offs in the metrics you proposed? How could you deal with striking a balance?

The operations manager wants to focus on minimizing the active number of vehicles to save costs while keeping an acceptable service level towards customers. Describe how would you help the manager out by using simulation techniques.

The company is judging a slow transition towards a different vehicle model that the one currently used. The main motivation is that the proposed new model is more reliable than the current one, meaning less breakdowns on the road. How could simulation be used to help take the decision of changing or not the models?

Optimization

A company builds motorbikes. The demand for the next months are: 110, 120, 130, 100. The initial stock is 20. Manufacturing a motorbike costs 1200€. The max capacity is 100 motorbikes per month but, with overtime labour, additional motorbikes can be manufactured at 1500€ each. Storing a motorbike in the warehouse for a month costs 100€. The warehouse can fit 25 motorbikes.

The goal is to satisfy demand while minimizing cost. Design an optimization problem to solve this. You need to include your target function, decision variables and constraints. Please, use clear notation that allows to understand the model.

Machine Learning

You work for a grocery delivery company. Customers build their grocery list through the company's app, select a day and hour for delivery and receive the goods at their home.

The company is growing by slowly adding new areas in the cities where it operates. Typically, the company grows one postcode at a time. Once a new postcode is opened up, the entire area contained in that postcode becomes available for the delivery service.

As the company has limited capital and resources, they need to pick carefully which postcodes to open up first. The underlying intuition is that not all postcodes are the same in terms of revenue and profitability, and the company will perform better if the "good" postcodes are included in the operations area first.

How do you think Machine Learning can be applied to tackle this challenge?

What data would you ask from the company? What data would you look for somewhere else? Why is the data you are picking relevant?

What would be your target variable? Would you model this as a classification or regression problem?

After some time working on this problem, your models don't seem to beat the baseline you set as a benchmark. What are your options in terms of next steps?

DRAFT

DRAFT

DRAFT