Applied Optimization and Machine Learning – 2021/2022

Final Exam – 30/06/2022

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

Simulation

1. Explain what simulation modeling is about.

2. Under which conditions or situations is simulation modeling a good tool choice?

3. Provide two drawbacks of simulation modeling.

4. Why is simulation modeling useful when studying complex systems?

5. Imagine you have to model an uncertain, real world phenomenon like demand for a product, but you don't have historical data that you can rely on. How would you decide how to simulate it?

6. What tools can be used to create a simulation?

3

7. What is simulation-based optimization?

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

9. What are heuristics and meta heuristics? How are they different? When would you use each?

10. During the course, we discussed how genetic algorithms can be adjusted to solve many different problems. What parts of the algorithm need to be tuned to each distinct case?

Optimization

1. What are the different parts of an optimization problem? What is their role?

2. How are optimization problems solved?

3. What would you do if your optimization problem does not have a feasible solution?

4. What is the issue with trade-off situations in optimization? Provide an example approach of how can trade-offs can be managed.

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

6. Why are piece-wise functions useful in optimization?

7. What is a solver?

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

9. What is sensitivity analysis?

10. Provide two circumstances where linear or integer programming would be a better tool than simulation.

6

Machine Learning

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

2. Why is labeled (as in, contains both X and y) data necessary for supervised Machine Learning?

3. What is the difference between classification and regression problems?

4. Why do we split data between train and test sets?

5. What defines a good split when building a decision tree?

6. Is letting a decision tree grow fully a good idea or not? Why? How would you decide what is the optimal size?

7. Why is accuracy typically not enough to measure performance in classification problems?

8. How can you know if a Machine Learning model is overfitting?

9. What would you do if your Machine Learning model is underfitting?

10. What is an ensemble method? Give two positive and two negative characteristics of ensemble methods compared to simple models.

Part 2

Simulation

You are working for an airport. The airport is designing a new terminal. The management needs to decide how many lines to put in the security control area (a line is composed of a metal detection arch + two conveyor belts, one at each side). Management is concerned about the purchase and operation costs of the rather expensive machines that need to be bought and maintained for each line that gets placed. But at the same time, they must make sure that the waiting times for passengers are bearable, avoiding long queues that make passengers unhappy and cause problems in the operations of the airport.

Describe how simulation can help the airport management in this situation.

What is the decision that needs to be made? What are the goals? How can the trade-off be managed?

What data would you ask for to design your simulation? How would you use it to model different parts of the reality in the simulation?

Describe (high level, no code) a heuristic to optimize the decision. You can assume you have access to a simulation code that allows you to run simulations of the security area in action.

Optimization

You work for a manufacturing plant. The plant develops three families of chemical products in three different production lines: cleaning agents, fertilizer and battery electrolites. All three lines require sulfuric acid in different proportions to work and manufacture their chemicals.

Each unit at each line produces the following revenue when sold:

- Cleaning agents: 10$/unit

- Fertilizers: 5$/unit

- Battery electrolites: 30$/unit

Each unit at each line requires the following amount of sulfuric acid to be produced:

- Cleaning agents: 0.4L/unit

- Fertilizers: 2L/unit

- Battery electrolites: 0.05L/unit

During the next year, the plant can acquire up to 300L of sulfuric acid at 2$/L. Beyond that, the company can acquire up to 2000L at 20$/L. The company must produce at least 250 units of fertilizer per year due to national regulation of the industry. The plant manager needs to make plans for next year on how much sulfuric acid to purchase and how to allocate it across the lines to get the best profits they can obtain.

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 the Marketing department of a car manufacturing company. The company is interested in building a ML model to classify tweets as negative or not negative towards the brand. A negative tweet is a tweet where the customer expresses being unhappy about the companies services or products. Luckily, the amount of negative tweets as part of the total population of tweets talking about the brand is pretty low (around 5%). But the company is hoping that such an automated classification algorithm can help their customer service agents focus easily on unhappy clients and their problems.

You can assume that the company can obtain all the tweets where the brand is mentioned from Twitter easily.

Propose a basic methodology to build the dataset to solve the companies request. This should, at least, include:

- What data would you propose using.

- Propose at least 5 features that you would build out of the data to prepare the training and testing dataset.

Decide on one performance metric and motivate your choice.

Propose a baseline algorithm for the problem.

After some work, you are pondering whether to build a simple decision tree classifier or go for a random forest classifier. Explain two advantages of each model type.

You finally decide to move forward with the random forest classifier. How would you decide what hyperparameters to use?

DRAFT

DRAFT

17

DRAFT