IEOR, IIT BOMBAY: PhD Admissions 2021 (December)

Test for shortlisted candidates

Read instructions below carefully

- You may refer to any material/book/notes for the exam. Kindly do NOT consult or collaborate with any person for this test.
- Answers to the questions of this test are to be submitted when you register for your interview. You may be asked to show your workings during the interview.
- There are **9** questions in this test.
- You can write your answers on plain sheets of paper. Each page should have your name AND application ID written on the top of the page. Question numbers should clearly written.
- To submit your answers, scan/take photo of your answer sheets and uploaded the files via the Registration Form.
- In case you write the answers using Word/Latex/etc, then you can upload the file via the Registration form.
- Note that questions from a wide range of topics have been asked. Attempt as many questions as you can. In depth answers to at least a few questions are expected.
- In case of a doubt about a question, you may make a reasonable assumption and clearly write it in your answer. Please do not contact the Admissions Committee for clarification on the questions.
- Your submission will be considered during the Interviews.
- **Deadline for registration for interviews is 11:59PM, December 05th, 2021, India Time.**

Questions

1. A medical facility for cancer patients deals with the following requirements.

   - Type 1 patients require investigation and examination and require to stay 1 day in the facility
   - Type 2 patients require a course of chemotherapy and require to stay in the facility for 5 days
   - Type 3 patients are being tried with a new treatment and require administration of the drug and observation, so require 2 days of stay and after 2 weeks, a further 2 days.

   The average arrivals of Type 1, Type 2 and Type 3 patients are n_1, n_2 and n_3 per week. You may use $n_1 = 3$, $n_2 = 4$ and $n_3 = 1$ to illustrate your answers.

   (a) For steady stream of arrivals (i.e. assuming exactly n_1, n_2 and n_3 referrals to the facility every week), how many beds are required to operate the facility?

   (b) Assume that beds are allotted on a First-Come-First-Served basis with no priority among customers of type 1, 2 and 3 (either within each type or across types).

   For random arrivals to the system (e.g. Poisson with known means n_1, n_2 and n_3), how would you estimate the bed requirement so that no more than k per cent of the arriving patients are turned away?

   (c) In addition, if type 1 patients stay between 1 and 2 days (equally likely) and type 2 patients stay between 4-6 days, how would your analysis change? **[10 Marks]**

2. Once admitted to the PhD programme at IIT Bombay, suppose you decide to cycle every day from your hostel to your lab. You may assume that every path is composed of several road-segments, each connecting one intersection to another. This may be visualized as a graph consisting of nodes (intersections) and edges (road-segments). Assume every road-segment goes either uphill or downhill, depending on the direction one takes. Your goal is to start from your hostel, go entirely uphill until some point, and then go entirely downhill to get to your lab! For example the path may consist of road-segments that go uphill-uphill-uphill-downhill-downhill, but not uphill-downhill-uphill. Assume that you have a campus map with m road-segments and n intersections, and that every road-segment takes a certain given amount of time to cover. Give a procedure to cycle from your hostel to lab meeting above requirements, and taking the smallest amount of time (as you also want to maximize research output while improving your fitness). **[10 Marks]**

3. The graph in Figure 1 shows the number of people entering and leaving a store (or Mall) over a 30-minute interval. The numerical scale of y-axis is NOT provided on purpose. You don't know exactly how many people were already there inside the place when you started observing.

   Graph the trend of the number of people in the store over the 30-minute interval.

   (Note: At time 0, indicates some initial number of people in the store by keeping a dot somewhere on the middle of the y-axis as shown in Figure 2 below) **[10 Marks]**

4. Write a program in any programming language to do the following:

   Take as input a series of processing time of jobs, and the due date of jobs. Then, the program should

   (a) Use Shortest Processing Time (SPT) rule to sequence the jobs. Output the job sequence, Makespan, Average flowtime, and Max Tardiness.

   (b) Use Earliest Due Date (EDD) rule to sequence the jobs. Output the job sequence, Makespan, Average flowtime, and Max Tardiness.

Figure 1: Number of people entering and leaving a store (or mall).

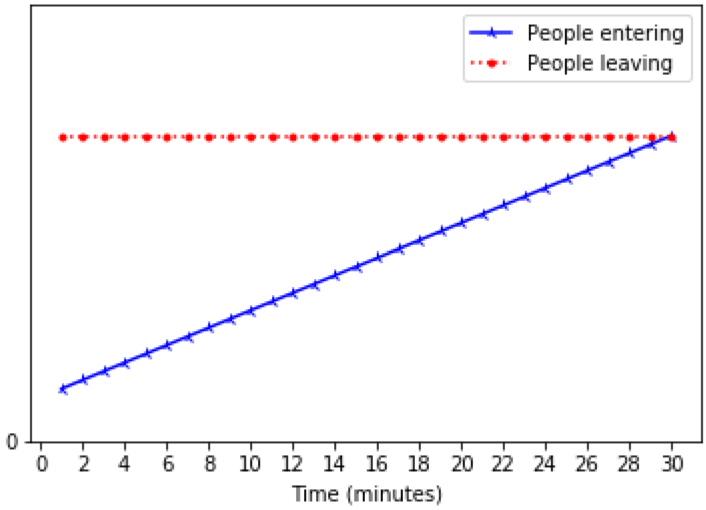
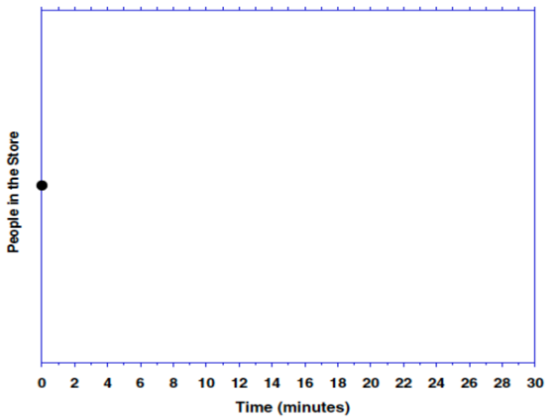


Figure 2: Number of people in the store.



Sample Test case (you can create more such scenarios to check):

Input (to be got as comma separated values only)	Output

`Process times: 11,29,31,1,2`
`Job Due Dates: 61,45,31,33,32`
	`SPT sequence: 4,5,1,2,3`
	`Makespan=74`
	`Avg flowtime=27.0`
	`Max tardiness=43`
	
	`EDD sequence: 3,5,4,2,1`
	`Makepsan=74`
	`Avg flowtime=47.0`
	`Max tardiness=18`

Note1: You need to submit the code as part of your submission (clearly indicate the programming language used).

Note2: The above scenario is called Single Machine Scheduling. You may view the Youtube video (from minute 30:27) to learn about Single Machine Scheduling/ SPT/ EDD, etc if required. **[10 Marks]**

5. The following questions are related to a Machine Learning task.

   (a) Assume a data set $\mathcal{D}$ containing one million Computerized Tomography (CT) scan images of human lungs. It is known that nearly 40% of these images contain details about lung tumors while the remaining 60% images depict lungs without tumors. However due to labeling effort, clinicians could only demarcate the tumor regions in a very few (around 100) of these images. Further, these tumor regions are marked using bounding rectangles. However note that only a few portions inside the rectangle correspond to tumor regions. Your aim is to develop a machine learning (ML) or deep learning (DL) algorithm to identify all the images in the data set (other than those already labeled) which contain the tumor regions and also to demarcate the tumor regions within these images using bounding rectangles. You need to provide details about the ML or DL technique you use and justify why it is useful for the task. If you use a DL technique, you would illustrate the network architecture of the deep neural network with details about the various components in the network and their purpose. You would also need to provide details about the loss function (or objective function) used when casting the ML or DL problem as an optimization problem. You would also provide the required details about the training algorithm, particularly emphasizing how your training algorithm will take care of the very few labeled images. Explain about the generalization capabilities of your training procedure.

(b) **Optional:** You can implement your training procedure and try it on some suitable open-source data sets. You can be ready with the code during the interview.

[10 Marks]

6. Consider the nonlinear optimization problem in two variables $\min x_1$ subject to $4x_1^2+x_2^2 \geq 4$ and $4x_1^2 + x_2^2 \leq 16$. Explain whether the point $(1, 0)$ is a point of a local minimum, local maximum or a point of stationarity. Find the globally optimal solution and check whether it satisfies the first order necessary (KKT) conditions for local optimality. **[10 Marks]**

7. Professor Viru comes into a class of 60 students, bringing a hat for each student with their roll number printed on it. Each student gets one randomly selected hat (among the 60 hats) by Professor Viru. Find the expected number of students that get their hats (The hat on which respective roll number is printed). **[5 Marks]**

8. In one huge room there are n-separate queues (waiting lines) and n- serving persons. One serving person serves exactly one of the waiting lines. Every queue has its own independent arrivals, these arrive according to an independent Poisson process with rate λ. At 10am there was an arrival to that room (to one of the queues). Let X represent the distribution of the time interval before the next arrival to the room (once again to any one of the queues). Find the distribution of X and explain properly. **[5 Marks]**

9. In Question 8 above, consider that the system starts at a time when each queue has at least one customer. All the n-servers start serving at exactly the same time. The job request of any customer requires 5 minutes to serve with probability 1/2 and 10 minutes to serve with probability 1/2. Let T represent the exact time at which all the n-servers finished the service of their respective first customers. Observe $T = \max_i T_i$, where T_i is the service time (time to finish the job request) of i-th server. Find the distribution of T. **[10 Marks]**

——————— End of questions ———————