

MG 265 Data Mining

Assignment 1

Due date: Tuesday, September 13 2022

1. An airline operates a set of flights daily with some fixed capacity. Each of these flights are called a flight leg. A passenger might use one or more of these flight legs to travel from his origin to reach the final destination, called the itinerary. The demand and cost of each itinerary is known. The problem faced by the airline is to determine the portion of demand of each itinerary to accept so as to maximise the total revenue. This problem can be formulated as a network flow optimisation problem.

Let the set of flight legs be \mathbf{L} and the itineraries be \mathbf{I} . Let the capacity of a flight leg $l \in \mathbf{L}$ be c_l . Also, let the fare paid by a passenger for an itinerary $i \in \mathbf{I}$ be f_i and the demand for the itinerary is d_i . The decision problem from the airline perspective is to determine the portion of demand x_i that should be carried for itinerary i without violating the capacity constraints and also maximise the total revenue generated. This linear programming problem is as given below in which the function $\mathcal{I}(l)$ maps a flight leg $l \in \mathbf{L}$ to all the itineraries using it.

$$\begin{array}{ll} \text{Maximise} & \sum_{i \in \mathbf{I}} f_i x_i \\ \text{Subject to} & x_i \leq d_i \quad \forall i \in \mathbf{I} \\ & \sum_{i: i \in \mathcal{I}(l)} x_i \leq c_l \quad \forall l \in \mathbf{L} \end{array}$$

Write a Python program using the `linprog` function in the `SciPy` library to solve this problem for the accompanying data set. Identify the flight leg that is capacity constrained.

In the flight leg capacity file (`cap.d`) you will find entries like 21 318 where 21 is the flight leg ID and 318 is the capacity of the flight leg. The itinerary flight leg mapping file (`it_legs.d`) has entries like 1 0 35, where 1 indicate the itinerary ID, 0 the first flight leg of the itinerary and 35 the second flight leg of the itinerary. In case of single flight leg itineraries the second flight leg will be marked with -1. The itinerary demand and fares by fare class (first line in these files) are provided in `it_fares.d` and `it_demand.d` files respectively. The itinerary ID + fare class constitute the members of the itinerary set \mathbf{I} .

2. Write a MapReduce program in Python for multiplying two matrices which is stored element wise (`A,i,j,a_ij`).