Homework 8

Given that we have sufficient information/data for each costumer x_i. This could be variety of factors such as age, how many people are in a household, what is the total income of that household, there past probability of paying for electric bill etc etc. By using logistic regression we can estimate the probability for x_i for the upcoming electric bill. So, if the output for x_i is 0.9, then household i is far more likely to not pay then a household x_j who has an output probability of 0.1. We can go further and try to classify probability intervals, for instance a household in the range of 0.9-0.99 is for more likely to not pay than a household that's in the range of 0.6-0.69. Before we do that however, we must first observe the distribution of probability (and make sure it fits a normal distribution) before we break it down to intervals . Then afterwards we break it down to ranges, then we can classify these ranges from 1-5, 1 being the highest likely to not pay and 5 being the most likely to pay. End result would be a grid map of the city with classified points ranging from 1-5. The next problem that arises is at which point does the electric company starts to turn off a household's electricity. This becomes an optimization question, where the goal is to find the point at which the electric company can shut electricity for a given household without losing too much cost for company , which is our constraint. After we found the sweet spot, say for example, from classification 4 to 5, the company can afford to shut off these household without too much damage to company's reputation and overall revenue . So let's assume for now that the cut off point is classification 4, and 5, we can imagine a grid map with points highlighting each classification point. So next we have the problem of logistics, how can we maximize efficiency of the drivers , whom are driving from house to house of a grid map. From here we can do a greedy approach, say driver starts at x_i who is classified as 5, after shutting off this household, it will take the next best option closest to that driver and heads that direction, so on and so forth. Again, this all depends on how frequent these points are on the map. Because if there are few(like 2 or 3), then we may have to re-evaluate everything and start from scratch, and maybe the 1-5 classification needs to be extended to 1-10 , this will ensure more points of the graph since the probability intervals are smaller. Through, trial and error I believe we can find the appropriate classification points that best maximizes the efficiency for the logistic team all while keeping cost for the electric company at minimum.