Project Proposal

Anna Midgley
annamidg@mit.edu

Nora Hallqvist
norahall@mit.edu

November 22, 2023

We are considering two different topics as we are unsure whether the first is a good idea or not. Please can you provide feedback on both, and whether you think either is better than the other.

1 Idea 1: Combing Covariate Shift with Stable Regression for Robust Model Training

1.1 Project summary

The proposed concept involves a two-step process that combines covariate shift adaptation with stable regression techniques to enhance model training. Our objective is twofold: firstly, to effectively mitigate dataset distribution differences, and secondly, to prioritize the training of challenging data points, all with the aim of improving model performance. The base model used for comparison will be a model which does not use stable regression. The benchmark for our evaluation will be a base model that does not incorporate stable regression and another one which does not include a covariate shift.

1.2 Methods

- Apply covariate shift: perform covariate shift by reweighting training data points to align them with the test set distribution. This is achieved by training a logistic regression model that predicts the source dataset for each data point.
- Apply stable regression: perform stable regression with the reweighted training data. This is done to train the model on the most challenging data points, enhancing its robustness. Models that will be trained: Linear regression, Logistic regression, OCT

Our project is an extension of stable regression that we learnt in class, where we seek to combine it with covariate shift. It additionally will employ three models that were taught in class.

1.3 Datasets

We plan to create a toy dataset that is synthetically created to have a covariate shift.

1.4 Challenges

This is a novel idea, so inherently there is uncertainty in whether the idea will work or not. A challenge could arise from the real-world datasets containing too great of a covariate shift, or too little of covariate shift.

2 Idea 2: Retail Store Inventory Optimization (from predictive to prescriptive)

2.1 Project summary

The project involves solving for inventory levels of different products for which revenue is maximized. The uncertain quantity of interest in the demand for each product type, the auxiliary data is attributes of the customer and product bought, and the decision variable is the inventory levels of each product. We will also explore different constraints placed on the decision variable for example, each product must be kept in stocked and the total volume of product stored cannot exceed a certain value.

2.2 Methods

Similarly to the homework, we will explore three methods for generating prescriptions. Specifically, 1) using sample average approximation, 2) using ML models for point prediction, 3) use ML models to predict weights. The ML models we will consider include random forests, OCT and OCT-Hs. SAA will serve as our baseline model.

2.3 Datasets

The dataset that we will use to predict customer demand is from Kaggle [1]. It contains various features related to customer shopping preferences such as age, gender and frequency of purchases.

2.4 Challenges

The main challenge will be coming up with realistic constraints in the inventory domain. For example, estimating how much volume each product will take up. We plan to make the problem harder, by adding more constraints that will ensure that the optimization problem is hard to solve.

References

[1] Kaggle. Customer shopping trends dataset, 2023. URL: https://www.kaggle.com/datasets/iamsouravbanerjee/customer-shopping-trends-dataset.