Exercise Sheet 4:

January 26, 2023

Probability Distributions and Causality

1) **Quantile regression**
   a) Predict the demand of all product-location-date combinations in `test.csv` in form of a full individual probability distributions (rather than mere point estimators, usually the mean of the underlying probability distributions, like in the exercises before) by using a method of your choice (quantile regression, generative method, or individually predict, e.g., mean and variance, under PDF assumption). It is fine if you predict several quantiles to approximate the probability distribution (for example by means of HistGradBoost from *scikit-learn*). You can choose one of the two setups described in exercise 2) a and b of exercise sheet 1.
   b) Evaluate your predictions for the 95th percentile by checking how close you are to the expectation of actual sales in the test data set being higher than the 95th percentile predictions in 5% of the cases. (An example for a use case is a subsequent order optimization with the goal to avoid out-of-stock situations in 95% of cases.)
2) **Qualitative and quantitative evaluation of PDF predictions** (See https://arxiv.org/abs/2009.07052 for a detailed description.)
   a) Plot a histogram of the CDF values of your predictions for the corresponding actuals.
   b) For a quantitative evaluation of your PDF predictions, calculate the earth mover's distance between your histogram from the last exercise and the expected uniform distribution.
3) **Causal Effects**
   a) Reducing prices or setting promotions is used for demand shaping, corresponding to an intervention in the causal language. Estimate the average causal effect (and, if you want, also the individual ones in terms of potential outcomes) of promotions on demand. This requires adjusting for all confounding variables, which need to be identified before.
   b) There are several products in the data set that, when in promotion, reduce the demand for other products in the same product group 3, an effect called cannibalization. Build a model (or a component of your overall model) to identify and predict cannibalization effects. You need to go beyond the usual i.i.d. assumption for this.