Homework Assignment 3

"Recommender System"

Course: Machine Learning in Marketing

Due date: 03.01.2021

In homework 02 you helped Dr. D to plan a marketing campaign for one of his hypermarkets by predicting which customers will not visit his store in the next two weeks. Dr. D told his friend Dr. S about how well the campaign went. Dr. S owns a small grocery store and is interested in your capabilities. He has launched a loyalty program and collected almost two years' worth of data. Now he is evaluating solution providers for the launch of his new couponing solution.

Just as most other retailers, Dr. S wants to run a pilot to test the quality of your predictions. The data for this test are available on Moodle:

`baskets.parquet`	Customers' past purchases (week, shopper, product, price in Euro Cents).
`actions.parquet`	Coupons customers received in the past (week, shopper, product, discount in percent).
`prediction_index.parquet`	See explanation below.

The file **`prediction_index.parquet`** serves as input for making your predictions. Your task is to predict the purchase ***probabilities*** for all customer-product combinations contained in the file. The time period for making predictions is week 90. With 2,000 shoppers and 250 products this results in 500,000 predictions.

Dr. S will evaluate your predictions using the metric ***log-loss***.

shopper	week	product	yhat
0	90	0	?
0	90	…	?
0	90	249	?
…	…	…	…
1999	90	0	?
1999	90	…	?
1999	90	249	?

Part 1   EDA

Before implementing your solution(s) to create purchase probability predictions, do some (meaningful) exploratory data analysis. For example, what is the true data generating process and how can you capture this in features? What information contributes to understanding customers' future purchases? How can you capture the impact of marketing variables and why is this important?

Part 2   Recommender system

Next, implement baselines. First build a predictor that outputs random probability predictions. Then implement one or two "management heuristics." Think about how managers might use simple descriptive statistics to determine future purchase probabilities.

Define a suitable test data set and benchmark the baselines' predictions.

Now it is time to implement your own model for predicting purchase probabilities. Benchmark your model and compare the results with the baselines.

Upload your solutions (probabilities `e03-$GROUP.parquet` on Moodle (Assignment 03) by noon, Jan 3. The leaderboard will be published in the exercise on Jan 7.

Bonus: Implement a stacked model: Produce multiple models that predict probabilities, then train an additional model that mixes the predictions (on a validation set).