Homework 1: Binary Choice

Due date: 8.11

The goal of this homework is to introduce you to google colab notebooks and simple binary choice problems.

1. Go over the google colab introduction notebook and see that you understand the basic components of google colaboratory.

2. Go over the introduction notebook provided and see that you understand the basic components of the training pipeline: user model, population model, training models and metrices. Make sure you get an accuracy higher than 80%.

3. Create the **SigmoidUserModel(UserModel)** class. Like the NoiseUserModel from the introduction notebook, this class models user valuations on items.

   The valuation of a user u on item x is modeled as follows:
   $$v(u, x) = sigmoid(u^T x) - \frac{1}{2}$$
   $$sigmoid(x) = \frac{1}{1 + e^{-x}}$$
   Also, create the SigmoidPopulationModel(PopulationModel) accordingly.

   Note: use np.random.randn(num_features) to generate user vectors u, similarly to NoiseUserModel.

4. Create a dataset with the generate_synth_data(num_features, num_items) function and the SigmoidPopulationModel class you created in section (3). Then run it through a training pipline with the accuracy metric and the LinearRegression() training model. Show the achieved accuracy.

5. Repeat section (4) but with a logistic regression training model.
   Note: replace LinearRegression() with MultiOutputClassifier(LogisticRegression()).

6. Run 100 iterations of sections (4)+(5). Meaning in each iterations generate a new dataset with SigmoidPopulationModel and train once with a linear regression model and once with a logistic regression model. Plot a graph of the accuracies in each iteration (2 graphs in the same figure – one for each training model accuracies).

7. Explain the results you got in section (6). Why do you think the graphs look like that?

8. Create the welfare(training_model, population_model, model_args, X) evaluation metric which will be used in our next workshop.
   The welfare of users U with respect to items M is defined as follows:
   $$Welfare = \frac{1}{|U|} \sum_{u \in U} \sum_{m \in M} \mathbb{I}\{sign(f(u, m)) = 1\} \cdot v(u, m)$$
   Where f(u, m) is the predicted valuation of the prediction model of user u on item m and v(u, m) is the true valuation of user u on item m. $\mathbb{I}$ is the indicator function
   $(I(x) = \begin{cases} 1, & x = True \\ 0, & x = False \end{cases})$.

9. Create the MSE(training_model, population_model, model_args, X) evaluation metric which will be used in our next workshop.
The MSE (mean squared error) of users U with respect to items M is defined as follows:

$$MSE = \frac{1}{|U||M|} \sum_{u \in U} \sum_{m \in M} \left(f(u,m) - v(u,m)\right)^2$$

Submission guidelines:

1. You can (if you want) create a copy of the introduction notebook and fill in your solutions at the end of it.
2. Submit a .ipynb file with your solution.
3. Add a text block in the beginning of your notebook with your IDs.
4. Indicate clearly with a text block the sections of your solutions
5. Answer dry questions in text blocks and wet questions in code blocks.
6. For any questions regarding this homework, contact Sagi.

Example for submission format:

IDs:

- 123456789
- 987654321

Question 1

```
[ ]  # My wet solution to question 1
```

Question 2

```
[ ]  # My wet solution to question 2
```

Maybe some dry explanations

Question 3

```
[ ]  # My wet solution to question 3
```

```
[ ]  # Continuing my wet solution to question 3
```