# Introduction

CA4015 Assignment 3/4

For our final assignment as part of the CA4015: Advanced Machine Learning module we were given the task of developing a recommender system. We would have to take an existing dataset and develop a recommendation system based off the most modern methods employed today. These methods were provided to us using an existing google colab book and it was up to us to implement and deploy these methods to our own data as we seen right. We also had to evaluate and look at our recommendations provided by the systems.
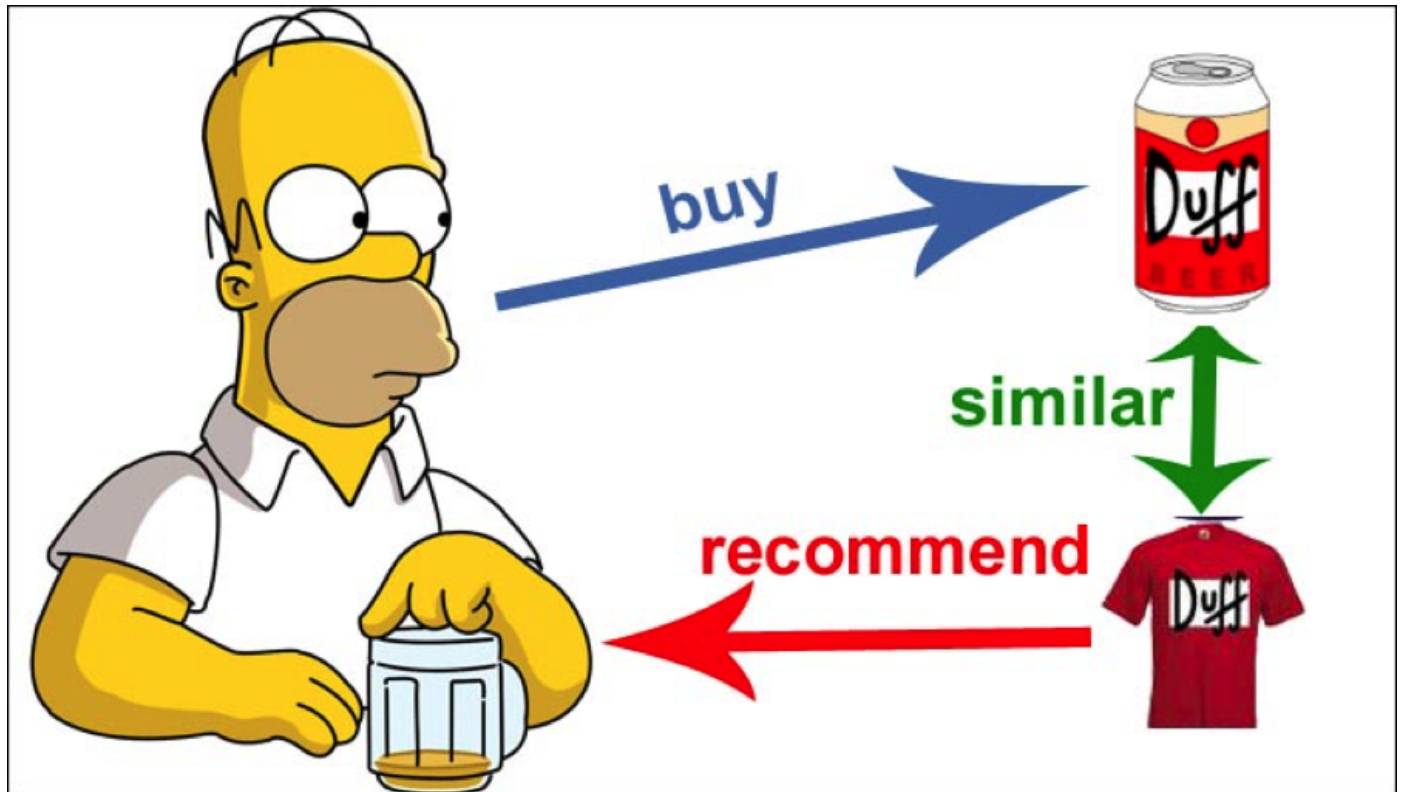
Dataset

We would be provided the lastFM dataset. This dataset consists of 5 files namely:

- artists.dat = information about music artists listened and tagged by the users in the data.
- tags.dat = information regarding available tags in the data.
- user_artists.dat = information abouut artists listened to by each user and listening count (weight) for each user/artist pair specified.
- user_taggedartists.dat = files contain the tag assignments of artists provided by each particular user and timestamp for each tag assigned.
- user_friends.dat = the users which are deemed as "friends".

The dataset was mostly clean but required a few minor fixes such as re-indexing among others. Some of the files related to the tags were problematic due to the inclusion of users not found in the other files related to user and artist ID's.

Outline of process

To start we will do some basic analysis of our data: what users listen to the most songs, what artists are most popular and so forth. Once this is done we will merge some of our dataframes together and begin to develop our models. We disregarded the "softmax" model in our process and implemented the regularized matrix model and basic model as per the colab provided. I wanted to compare this model to something else and this is what I did, comparing it to a system which utilises a neural network and works with feature embeddings. Finally, I tested our regularized model on my own spotify account to see what artist recommendations it could make to me based on my favourite artists. We also clustered the data on the top 100 tags and the mean amount of times artists related to these tags were played.

Links

- [Git repo](#)