Recommendation System

Renner Lucena - renner@stanford.edu

Predicting

Recommendation systems are a classic problem of AI. However, the specific challenges that show up while creating them, looking at a specific industry, company, and the data available, make it really interesting to build one. In this class, I did it for a startup that sells tickets online. Since the data of users was limited and restricted, even given privacy issues, the model now basically just uses the data of events. It predicts, with a little more of 70% of accuracy, if a certain event will be successful or not, a label that was given to me by the company based on a complex and strategic query that measures the satisfaction of users and also how advantageous was the event for the company. The input for it were features of the event.

Data

The features of an event taken in consideration were month, day_of_month, day_of_week, fractional_hours (time), latitude, longitude, category, average_price (of the ticket), ticket_amount (number of tickets released for sale), and the label was 1 or 0, if the event was successful or not.

Features

The basic ones were already explained previously. Since these represent relevant data for the model directly, nothing specific seemed to be derived from them, only the formation of "clusters" around some locations and times. They seem obviously appropriate because these features describe the event.

Model

Searching online, and talking to advisors, I could not find an architecture that was specifically good for my task. Also, there were not heuristics to help me with the decision of an architecture. After a lot of iterations though, I got a NN with 10 layers,, with dimensions 45, 40, 35, … 5, (multiples of 5, starting from the number of features, 9), all RELU, followed by Sigmoid in the end.

Results

In total, 3613 samples of events were provided with labels from the company's database, 3300 for training and 313 for testing.

Training	Testing
~0.86	~0.71

Discussion

I think that the project was nice, and I do think that I learned a lot, and that it was pretty rewarding to see things working. For sure, I wanted better results, but 70% is already pretty good. I was not even sure that it was achievable, given the limited amount of data available, and the uncertainty around how the features and the label given were correlated.

Future

I would try to get more data, test some more complex architectures for the NN, see if I could find heuristics about the shape of the layers, and possibly add more features extracted from the data of the company.

References were omitted since there were not papers found related to this specific task. Only class material was used.