CSE 515T Project Proposal

A Fully-Bayesian Artificial Neural Network for Churn Prediction

By Patrick Di Rita

Problem and Motivation

Institutions such as banks, insurance providers, and utility companies are highly incentivized to predict whether they will lose a customer. The reason for this is simple: it is much more expensive to sign a new client than to keep an existing one. The loss of clients or customers is known as customer churn, and many companies have invested a great deal of resources into understanding not only which customers are at a high risk of leaving, but also the causal factors of churn. As companies' ability to collect customer data has improved, the use of modern predictive analytics for churn modeling has become extremely common across a multitude of industries.

The most notable such industry is banking, as banks structure their entire business models around customer retention, and it can be quite easy for a customer to simply switch banks (as opposed to the cable television industry, where many companies hold a near monopoly for some regions). Banks require their churn models to be not only extremely accurate, but to also be both simple and fast, as many banks work with massive amounts of data and cannot afford high model training times. As such, recent years have seen many banks using artificial neural networks (ANNs) for churn modeling, as customer data is easily represented in tabular form.

The issue with ANNs (and neural networks in general) is that many engineers design complex neural network architectures for problems that might not require high complexity, and the architectures are never simplified, as the benefit does not seem to outweigh the cost of redesigning, re-testing, and re-deploying a new model. We aim to remedy this issue at the source, by using Bayesian optimization to build an ANN from scratch. Bayesian optimization has been shown to be very good at finding combinations of parameters that engineers would usually not have considered, and this can effectively be applied to constructing optimized neural network architectures.

Our Modeling Plan

We plan to build an ANN for predicting banking customer churn by maximizing the model's predictive accuracy as a function of the neural network's architectural hyperparameters. We will use Bayesian optimization with a Gaussian process model to discover optimal selections of the number of layers and their dimensions, amount of dropout, and activation functions, as well as more general hyperparameters such as optimal early stopping, optimizer hyperparameters, batch size, and number of training epochs. We will also add a term to the accuracy calculation that penalizes high model complexity, in order to encourage exploration of the area of the search space containing simpler parameterizations.

We also plan to add additional automation by implementing a pooled acquisition function strategy. More specifically, we will select a parameterization for observation by having an ensemble of acquisition functions each nominate a point, and select a nominated point with a probability corresponding to how

successful the point's acquisition function has been at giving good nominations in previous trials. This essentially turns the acquisition function step into an additional multi-armed bandit problem.

The overall goal is to design a system which can generate a simple yet effective ANN without requiring expert input for the model architecture and acquisition function steps. In theory, a bank would be able to deploy this system to their analytical team, and team members could generate models and accurately predict customer churn without requiring oversight by a machine learning engineer.

Data

The dataset contains information on 10,000 customers at an anonymized European bank. For each customer, there are data fields such as age, gender, home country, credit score, years at the bank, etc., along with a binary target variable reflecting whether the customer closed their account at the bank or continues to be a customer.