

For the last program of the semester, we'll use TensorFlow to build a neural network to analyze some real-world data.

Telecom providers measure their customer retention by *churn*, the percentage of customers who leave the company in a given time period. While this will never reach 0 and a company probably shouldn't try for 0 churn, a high level of churn can indicate customer service problems. And since recruiting new customers is expensive, much more expensive than retaining existing customers, a high level of churn can become an expensive problem.

This dataset, downloaded from [kaggle.com](https://www.kaggle.com/wmccomas), was compiled by IBM. A quick summary of the data from that site:

Each row represents a customer, each column contains customer's attributes described on the column Metadata.

The data set includes information about:

- Customers who left within the last month – the column is called Churn
- Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies
- Customer account information – how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges
- Demographic info about customers – gender, age range, and if they have partners and dependents

A more detailed breakdown is given at the end of this document.

Your task is to develop a neural network that will predict (classify) whether a particular customer will churn or not.

Development notes:

- There is no clear optimum number of layers, or size of layers. Experiment and see what happens as you expand, reduce, or add layers.
- Some data is categorical; there are 2 or 3 non numeric values. For example, the type of contract may be "month-to-month," "one year", or "two years". These should probably be coded using 1-hot coding. Whether to normalize numeric values or leave them alone is something you may want to explore.
- Express your final estimate as a probability using softmax.
- As a rule of thumb, use 70% data for training, 15% for testing, and 15% for validation.
- Your program should read "program4input.txt" from the default directory.

Write a short report (2 or 3 pages) describing your network and the results you get. Submit this report along with your TF source file.

Appendix: Data dictionary

customerID: Customer ID

gender: Customer gender (female, male)

SeniorCitizen: Whether the customer is a senior citizen or not (1, 0)

Partner: Whether the customer has a partner or not (Yes, No)

Dependents: Whether the customer has dependents or not (Yes, No)

tenure: Number of months the customer has stayed with the company

PhoneService: Whether the customer has a phone service or not (Yes, No)

MultipleLines: Whether the customer has multiple lines or not (Yes, No, No phone service)

InternetService: Customer's internet service provider (DSL, Fiber optic, No)

OnlineSecurity: Whether the customer has online security or not (Yes, No, No internet service)

OnlineBackup: Whether the customer has online backup or not (Yes, No, No internet service)

DeviceProtection: Whether the customer has device protection or not (Yes, No, No internet service)

TechSupport: Whether the customer has tech support or not (Yes, No, No internet service)

StreamingTV: Whether the customer has streaming TV or not (Yes, No, No internet service)

StreamingMovies: Whether the customer has streaming movies or not (Yes, No, No internet service)

Contract: The contract term of the customer (Month-to-month, One year, Two year)

PaperlessBilling: Whether the customer has paperless billing or not (Yes, No)

PaymentMethod: The customer's payment method (Electronic check, Mailed check, Bank transfer (automatic), Credit card (automatic))

MonthlyCharges: The amount charged to the customer monthly

TotalCharges: The total amount charged to the customer

Churn: Whether the customer churned or not (Yes or No)