

Team Members:  
Jasjeet Dhaliwal  
Aaron Giera  
Ekin Koker

Project Title: Healthcare Events Prediction Using Neural Networks

Our dataset, collected by fellow grad student Gabriel Schmitt from Medical Expenditure Panel Survey database (<https://meps.ahrq.gov/mepsweb/>), consists of demographic data, healthcare data and healthcare events of 58801 patients over the course of 5 years, between 2010-2014. The patients are surveyed for two years, so both data and events are recorded for two years. Healthcare data is expressed in Clinical Classification Codes (CCCs) and show whether a patient has a particular health condition or not. This data is very sparse since there is a high number of CCCs (670 in our data, to be exact) and most patients have a few number of these conditions. There is also demographic data of patients specified in 34 dimensions. Finally, the healthcare events are expressed in specialist (office) and hospital visits, specified in 40 dimensions.

We are interested in predicting these healthcare events using neural networks (other machine learning methods are also of interest). We are specifically interested in exploring the change of events from the first year to second year. To keep the task manageable, we are considering limiting ourselves to learning first year events using NNs to predict second year events, but if necessary, the demographic and health care data will also be incorporated. Dimensionality reduction methods might be used to reduce dataset into a more manageable size. Another interesting aspect of the data could be correlations between demographics and health conditions.

Project Title: Learning and Replicating Music through MIDI files Using GANs

MIDI files are a specific types of music format that carries event messages that specify notation, pitch and velocity and they have a very distinctive advantage which is compactness - an entire song can be coded in a few hundred lines, i.e. in a few kilobytes. (Information on MIDI files resourced from Wikipedia - [https://en.wikipedia.org/wiki/MIDI#Standard\\_MIDI\\_files](https://en.wikipedia.org/wiki/MIDI#Standard_MIDI_files)) We are interested in collecting a MIDI dataset consisting of different composers and teaching a GAN to identify and replicate the work of these composers. TensorFlow's Magenta MIDI Interface (<https://magenta.tensorflow.org/2016/08/02/midi-interface/>) might be of use. If the Generator module of GAN can create MIDI files that Discriminator module cannot separate from a real composer's MIDI file, we would consider our project to be successful. If our Discriminator module is poor in comparison with our Generator module we could meet this condition while our project does not perform well. Music is a subjective experience, so part of our measure of successfulness must be our direct opinion of some amount of samples. We could also ask some friends or colleagues to try and distinguish a few random samples of music as human or machine composed.