Hospital Readmission Prediction of Patients using Deep Neural Networks

Chahes Chopra (IPG-2014031)
Shivam Sinha (IPG-2014082)
Shubham Jaroli (IPG-2014085)

June 30, 2017

Group No - 23

Supervisor Name - Prof. Anupam Shukla

1 Problem Statement

To predict whether a patient will be readmitted or not using a dataset of 70,000 Clinical Database Records with Deep Neural Networks.

2 Methodology

The first task is to preprocess the raw dataset.After the preprocessing, we will be using Deep Neural Network (DNN) to predict patients readmitted. In the base paper, the impact of HbA1c measurement was used to predict readmission rate using multivariate logistic regression but we will advance this and use DNN to predict whether a patient will be readmitted or not on several factors and not only on HbA1c test.

3 Experimental Setup (Dataset, Software, Hardware Used)

Data Set Information:-
The dataset represents 10 years (1999-2008) of clinical care at 130 US hospitals and integrated delivery networks. It includes over 50 features representing patient and hospital outcomes. Information was extracted from the database for encounters that satisfied the following criteria.
(1)It is an inpatient encounter (a hospital admission).
(2)The length of stay was at least 1 day and at most 14 days.

(3)Laboratory tests were performed during the encounter.
(4)Medications were administered during the encounter.
The data contains such attributes as patient number, race, gender, age, admission type, time in hospital, medical specialty of admitting physician, number of lab test performed, HbA1c test result, diagnosis, number of medication, number of outpatient, inpatient, and emergency visits in the year before the hospitalization, etc.

For the coding, we will be using Python(version 2.7) and Anaconda platform and its libraries like -
Numpy - For matrix computations in DNN
Pandas - For matrix computations in DNN
Tensorflow - An open-source software library for Machine Intelligence..
Keras - Keras is a high-level neural networks API, capable of running on top of either TensorFlow, CNTK or Theano.
Sklearn - A open source library for data mining and machine learning.
Random - To generate ranodom numbers used
Matplotlib - To create graph,charts and other visualisations of our model.
SciPy - To read our matrix dataset.
Math - For the basic mathematics operations.

We will be using our personal laptops for the project work. The computer on which we are developing our model has 8GB of RAM, intel core i7 processor(2.7 GHz octacore) and NVIDIA GeForce 820M graphic card.