Bowen Tan, Bo Wu, Yuetong Li

Introduction

• Describe what you are planning to do.
We plan to use the physical condition and living habits of a person to predict the probability of getting a stroke through a machine learning model.

• What would the successful outcome of your project
We input a person’s data of his condition as parameters, such as age,gender, and smoking status, to predict the person’s probability of getting stroke.

• Describe why your project is interesting. Describe the motivation from a personal learning perspective.
According to the World Health Organization (WHO) stroke is the 2nd leading cause of death globally, responsible for approximately 11% of total deaths. We hope our project will remind people to take precautions against the stroke. A stroke is dangerous but still preventable with a healthy lifestyle and medicine. So we try to find the connection between a person's physical condition and the probability of getting a stroke by using ML.

Analysis plan

• Here you describe the ML models to be used

We choose to use a polynomial regression model, which is a type of linear regression model. Here the model is in the abstract form of y=w0*x0+w1*x1+w2*x^2+...+wn*x^n, where (w0, w1, …wn) are parameters of each variable (x0, x1, ...xn). We will train this model based on the stroke dataset we got and use the trained model to predict the probability of a person gets stroke.

• Describe your data set, including the data source link.

The dataset we choose is a stroke dataset with 5110 observations and 12 variables, including 7 variables with numerical type and 5 with string type storing in a csv file. This file collected 5110 people’s body and living habits, such as bmi, age, marcial and smoking condition, with the last column containing the result whether each person troubles from stroke or not.
The source of our dataset comes from: https://www.kaggle.com/fedesoriano/stroke-prediction-dataset