An Analysis on Public Health Data of Chronic Kidney Disease

Akshay Sanjeev

July 9, 2025

1 Motivation

1.1 Notation and Convention

By standard dataset, from now on I will be referring to the following data structure. We have m independent variables(a.k.a features), denoted by $x^i, i = 1, 2, 3, \cdots, m$ and one dependent variable(a.k.a label), m. Let n be the total number of observations, and the lower index denotes the observation. Thus the dataset is set like the one given below.

$$\begin{pmatrix} x_1^1 & x_1^2 & \cdots & x_1^m & y_1 \\ x_2^1 & x_2^2 & \cdots & x_2^m & y_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_n^1 & x_n^2 & \cdots & x_n^m & y_n \end{pmatrix} \tag{1}$$

2 Pre-Processing Data

3 Decision Tree

3.1 Informal Description

A decision tree is a supervised learning method, where you split data based on a nested `if else` conditions. To build a decision tree, take the standard dataset,

4 Random Forest

4.1 Informal Description

5 Feed-forward Neural Network

5.1 Informal Description