1. 目的、研究動機

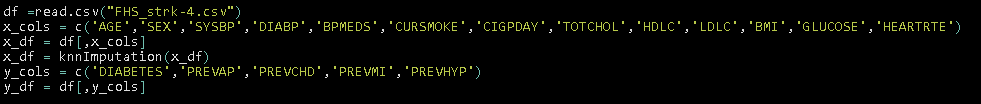
試圖以PLS手法，分析此份資料，找出對DIABETES、PREVAP、PREVCHD、PREVMI、PREVHYP等五種症狀(以下簡稱目標疾病)關鍵影響因子，若有複數，則將複數因子轉化成綜合評估指標，試圖以此指標評估得到人群是否為高風險群。

1. 手法: KNN、PLS、PCA、ANOVA
2. 資料分析
3. 資料前處理

X資料集 : AGE,SEX,SYSBP,DIABP,BPMEDS,CURSMOKE,CIGPDAY,TOTCHOL,HDLC,LDLC,BMI,GLUCOSE,HEARTRTE

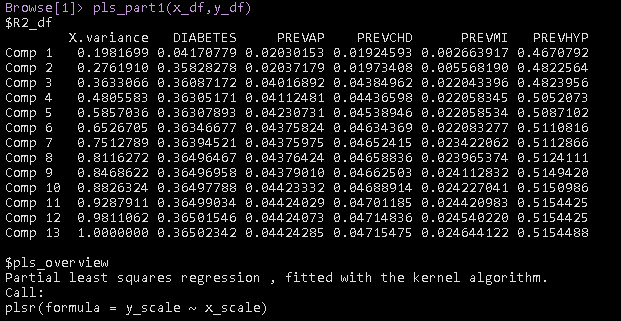
Y資料集 : 目標疾病

將X資料集缺值部分利用KNN手法進行補值

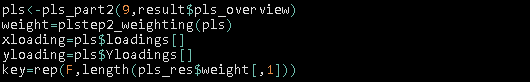


1. PLS分析

-找出X因子對Y因子解釋能力較佳的選項後，進行分析

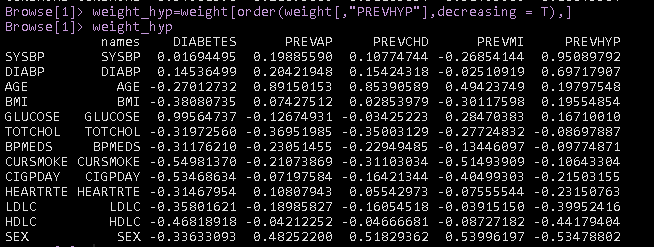


由數據可以看出，只有DIABETES、PREVHYP兩者解釋能較佳，在不遺失過多X資訊下，挑選Comp 9較佳。

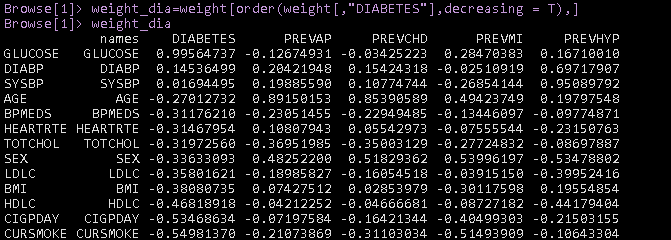


得到權重後，再分別以DIABETES、PREVHYP進行權重排序

PREHVP



DIABETES



以正值作為挑選原則，

PREHVP 關鍵因子 : SYSBP、DIABP、AGE、BMI、GLUCOSE

DIABETES 關鍵因子 : SYSBP、DIABP、GLUCOSE

挑選PREHVP做為後續分析項目，

3.PCA資料降維