Assignment
Supervised learning

Diabetes data set

Context

This dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases. The objective of the dataset is to diagnostically predict whether or not a patient has diabetes, based on certain diagnostic measurements included in the dataset. Several constraints were placed on the selection of these instances from a larger database. In particular, all patients here are females at least 21 years old of Pima Indian heritage.

Content

The datasets consist of several medical predictor variables and one target variable, Outcome. Predictor variables includes the number of pregnancies the patient has had, their BMI, insulin level, age, and so on.

Columns

Pregnancies Number of times pregnant
Glucose Plasma glucose concentration a 2 hours in an oral glucose tolerance test
BloodPressure Diastolic blood pressure (mm Hg)
SkinThickness Triceps skin fold thickness (mm)
Insulin 2-Hour serum insulin (mu U/ml)
BMIBody mass index (weight in kg/(height in m)^2)
DiabetesPedigree FunctionDiabetes pedigree function
Age Age (years)
Outcome Class variable (0 or 1)

Perform below using supervised techniques you learnt.

- Outcome variable is a class type variable with two values 0 and 1.Build KNN for range of values of K i.e 1....50 and identify the optimum value of K for maximum accuracy.