Assignment - A2 41403

- Title: Naive bayes classification
- Problem statement: Download PIMA Indians Diabetes dataset. Use naive bayes algorithm for classification. Load the data from csv file & split it into training & test datasets. summarize the properties in the training dataset so that we can calculate probabilities & make predictions classify samples from atest dataset.
- Objective: Understand Naive bayes algorithm for classification.
- Outcome: predict whether the person has diabetes or not using Naive bayes classifi based on parameters in dataset like blood pressure.
- S/W H/W requirements: 64 bit OS (UNIX/ LINUX) python3, Jyputer environment.

- Theory: Naive bayes classifiers are a family of simple probalistics classifier They are based on baye's theorem which describes the probability of a certain event occuring based on the prior knowledge of conditions that might be related to event

bayes theorm is stated mathematically

$$P(A|B) = \frac{P(B/A)\ P(A)}{P(B)}$$

41403

where, A·B are the events.
P(A|B) is a conditional prob. The likelihood of event A occurring knowing that B is true
P(B|A) is a also conditional the likelihood of B occurring knowing that A is true
P(A) & P(B) are marginal probabilities

Naive bayes is a technique for constructing classifiers which applies the above theorem with the strong assumption
These models assign class labels to problem instances. represented as vectors of features values

A family of algorithms based on one common principle from the naive bayes classifier the principle is that,
a particular feature is independent of the value of any other feature. given the class variable each feature contributes independently to the probability of the positive outcome.

- About the data set:
The dataset is originally from the National institute of Diabetes & Digestive & kidney Disease.

41403

The objective of the dataset is to diagonostically predict whether or not patient has diabetes based on certain diognostic measures included.

Several constraints were placed on the selection of these instances from the larger database in particular all patients here are at least 21 years old.

- conclusion: The Naive Bayes classifier was successfully applied to the cleaned dataset & the out come was predicted with an good accuracy.