Rohit Kulkarni 41346

M T W T F S S
Page No.:
Date:
YOUVA

DA ASSIGNMENT 2

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. Summarise the properties in training dataset so that we can calculate probabilities & make predictions. Classify samples from a test dataset & a summarized training dataset.

OBJECTIVES: Understand Naive Bayes Algorithm for classification & use it on Pima Indians dataset.

OUTCOME: Predict whether the person has diabetes or not, using Naive Bayes Classification based on the parameters in dataset.

S/W AND H/W REQUIREMENTS:

64-Bit OS, python, jupyter notebook, keyboard, mouse, monitor.

THEORY:

i) Naive Bayes classifiers are a family of simple probabilistic classifiers.

ii) They are based on Bayes theorem, which describes the probability of a certain event occuring, based on the prior knowledge of conditions that might be related to the event.

Bayes theorem is stated mathematically -

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

where A, B are events.

P(A|B) is the probability the likelyhood of event A occurring, knowing that B is true.

P(B|A) is also conditional probability, the likelihood of B occurring knowing the A is true.

P(A) & P(B) are marginal probabilities.

i] Naive Bayes is a technique for constructing classifiers which applies the above theorem, with the strong assumption that the features are largely independent.

ii] These models assign class labels (Diabetic or Non-diabetic) to problem instances, represented as vectors of feature values.

iii] The class labels are drawn from a finite set.

The principle is given as -

'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, regardless of any possible correlations between the features.

M T W T F S S

Page No.:

Date:

YOUVA

i] Abstractly, Naive Bayes is a conditional probability model and can be trained very efficiently between the features.

ii] Despite its naive design & apparently oversimplified assumptions Naive Bayes classifiers have proven to work quite well in real world settings.

CONCLUSION:

The Naive Bayes Classifier was successfully applied to the Pima dataset & the outcome (Diabetes diagnosis) was predicted successfully.