

MACHINE LEARNING: ASSIGNMENT

Naïve Bayes Classifier

Group Members:

Michael Nyembe: 1278687

Refentse Bane : 1390445

Motheo Sekgetho: 1432706

>

Data Description - This database encodes the complete set of possible board configurations at the end of tic-tac-toe games, where "x" is assumed to have played first. The target concept is "win for x". "x" has one of 8 possible ways to create a "three-in-a-row".

Attribute Information:

- top-left-square: {x,o,b}
- top-middle-square: {x,o,b}
- top-right-square: {x,o,b}
- middle-left-square: {x,o,b}
- middle-middle-square: {x,o,b}
- middle-right-square: {x,o,b}
- bottom-left-square: {x,o,b}
- bottom-middle-square: {x,o,b}
- bottom-right-square: {x,o,b}
- Class: {positive,negative}

What we are trying to predict: A win for x

There are 958 complete datapoints. For Naïve Bayes, the data was split into 574 training data points and 384 testing data points.

Some examples of data points:

- x,b,o,b,o,x,x,x,positive
- b,x,x,o,x,b,o,x,o,positive
- o,x,b,o,o,x,o,x,x,negative

>

Processing Inputs/targets

- The inputs/targets for this classifier were pre-processed using the *preprocessing* method from *sklearn*. This converted the x's to 2, o's to 1 and b's to 0 for the inputs, and it converted the positives to 1 and negatives to 0 for the target/class.

Splitting of data

- The *train-test-split* method from *sklearn* was then used to split the data into training and testing data. The data was split at a 0.4 ratio and randomly sampled.

>

- The Standard Naïve Bayes was used to classify the data without any smoothing applied to the algorithm since the input data was enough hence it wouldn't yield any output of 0 for any entry.
- The accuracy of classification increases when using the built in Gaussian Naïve Bayes. It yields an accuracy of 0.72396 which is the highest attainable accuracy value for this dataset when using naïve bayes.
- Algorithms such as Instance Based and k-Nearest Neighbour are believed to work better on this dataset according to research and experiment performed on this dataset.