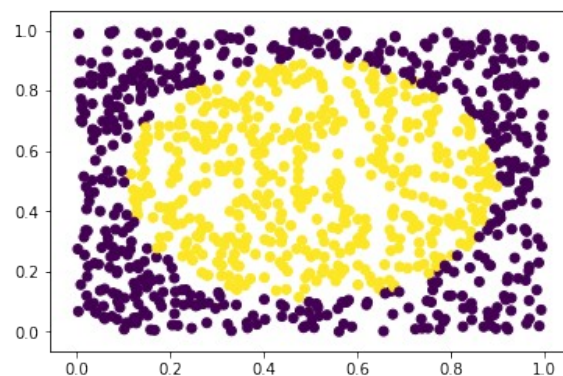**Test**

In this section we provide the description of the used dataset and of the network built through our framework. Then, the validity of the model is shown by comparing it with well-known baselines.

Dataset

The dataset consists of a training set and a test set of 1000 points each. These points are sampled uniformly in $[0, 2]^2$, and they are labeled with 0 if they end up out of the disk of radius 1/sqrt(2*pi) (label is 1 otherwise). The following exhibit – which represents the train set used during the testing phase – contains a set of points that satisfy these properties:

For the sake of reproducibility, we used a manual seed to ensure that all the runs of the test would be the same.

Network

The network consists of two input units, two output units and three hidden layers of 25 units.

A Tanh activation function is used as the last component of our network. The reason behind this choice is that since we are using MSE in a classification problem, having an unbounded output would highly penalize the outcome. In fact, while the output can assume any value, the labels can only have two values. This leads to the following problem: any value that is either too high or too small is seen, by the MSE function, as something very different from the labels.

By using the Tanh this problem is mitigated and any output value is approximated to the closest label, if its absolute value is too high.

Model Comparison

The accuracy of our model was tested with the aforementioned dataset. The results were quite satisfying, since the accuracy is XX +/- YY.

Then, a comparison with common baselines was performed:
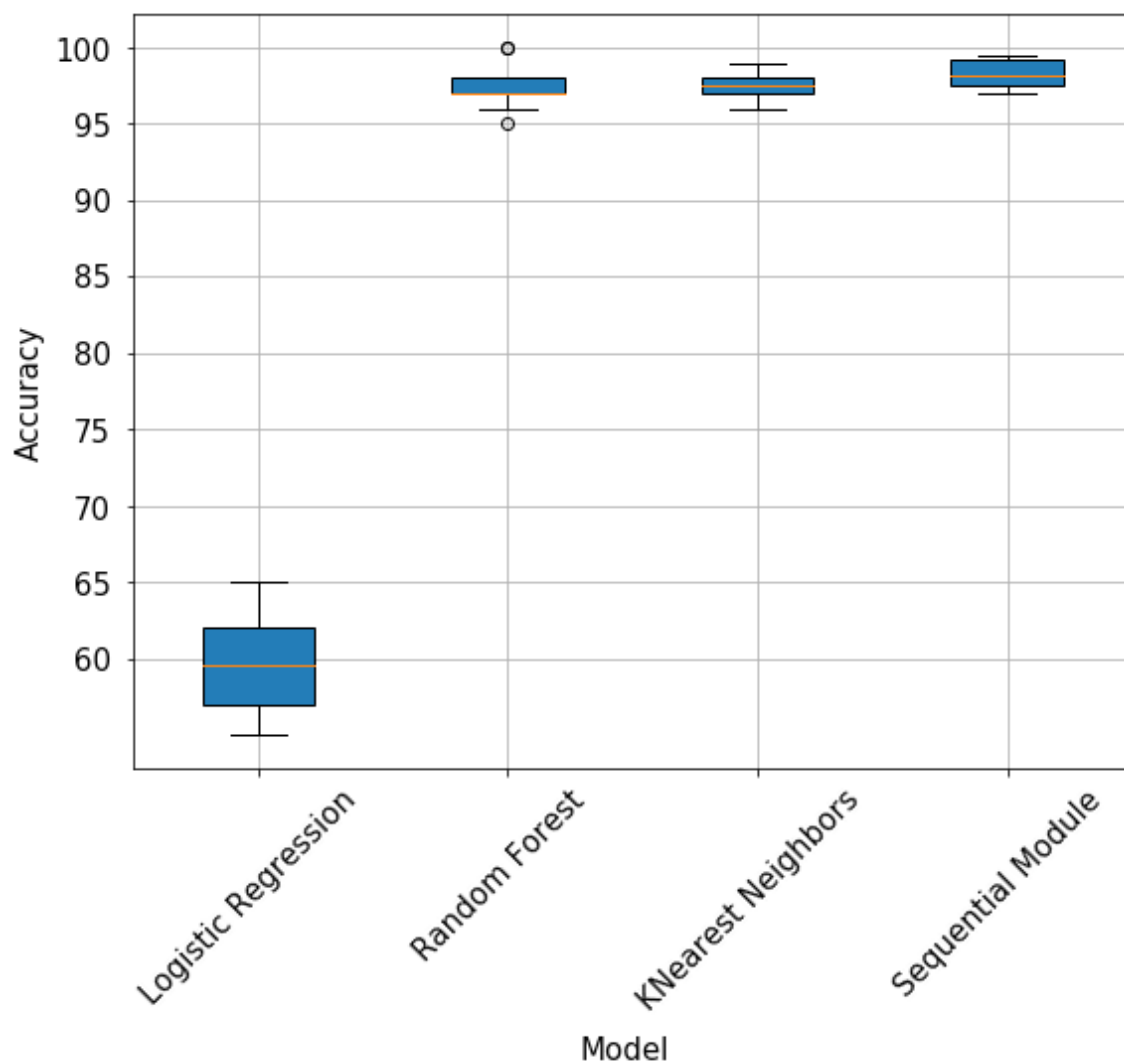
As the above exhibit shows, non-linear classifiers (K-NN, random forest) perform reasonably well, while linear ones (logistic regression) give the worst result because the points in the dataset are not linearly separable. In this latter case, better performance could have been achieved through feature extraction.

It can also be seen that our model performs slightly better than any of the tested baselines.