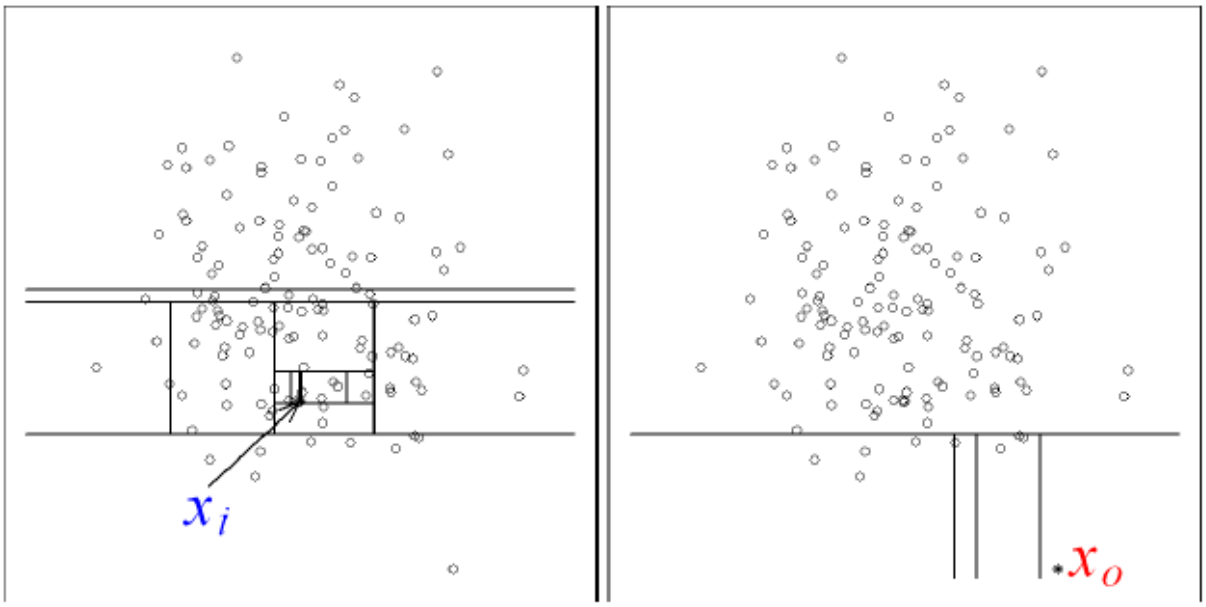
Assignment 7

If a tree falls in a forest… - Isolation Forests



(a) Isolating x_i (b) Isolating x_o

In this assignment you write code “from scratch” to implement an isolation forest, and then test it on a provided dataset (based on MNIST-C).

Specifically, you will:

- Get the provided dataset from D2L.
- Write your own version of isolation forest code.
- Run your code on the dataset to obtain anomalousness scores for each point for your chosen parameter settings.
- Sort the points by anomalousness scores and generate a precision-recall curve.
- Generate the equivalent of the following figure from your forest.

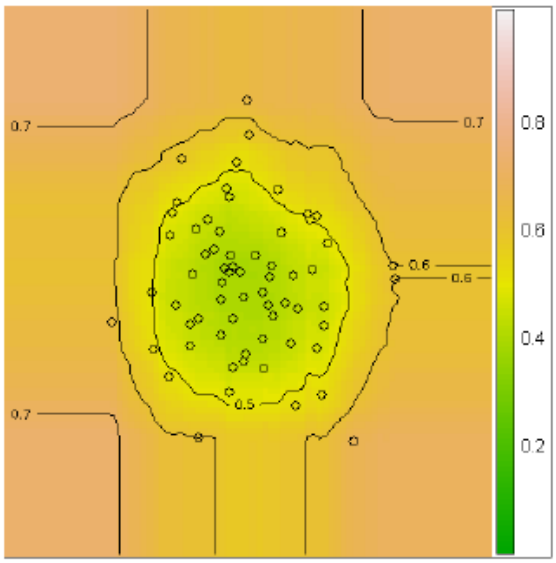


Fig. 5. Anomaly score contour of *i*Forest for a Gaussian distribution of sixty-four points. Contour lines for $s = 0.5, 0.6, 0.7$ are illustrated. Potential anomalies can be identified as points where $s > 0.6$.

You will turn in both the code and a short report, discussing what you've implemented, how well it worked, what you've learned, etc.

Remember, you must understand what you turn in – you may be asked to explain it to me and/or the class.