Dataset information

The original [lymphography](https://archive.ics.uci.edu/ml/datasets/Lymphography) dataset from [UCI machine learning repository](https://archive.ics.uci.edu/ml/) is a classification dataset. It is a multi-class dataset having four classes, but two of them are quite small (2 and 4 data records). Therefore, those two small classes are merged and considered as outliers compared to other two large classes (81 and 61 data records).

Source (citation)

A. Lazarevic and V. Kumar, “[Feature bagging for outlier detection.](http://neuro.bstu.by/ai/To-dom/My_research/Papers-0/For-research/D-mining/Anomaly-D/KDD-cup-99/p157.pdf)” in ACM SIGKDD, 2005, pp. 157–166.

A. Zimek, M. Gaudet, R. J. Campello, and J. Sander, “[Subsampling for efficient and effective unsupervised outlier detection ensembles.](http://www.dbs.ifi.lmu.de/~zimek/publications/KDD2013/subsampling-outlier-ensemble.pdf)” in ACM SIGKDD, 2013, pp. 428–436.

C. C. Aggarwal and S. Sathe, “[Theoretical foundations and algorithms for outlier ensembles.](http://www.kdd.org/exploration_files/Article4.pdf)” ACM SIGKDD Explorations Newsletter, vol. 17, no. 1, pp. 24–47, 2015.Downloads

Downloads

File: [lympho.mat](https://www.dropbox.com/s/ag469ssk0lmctco/lympho.mat?dl=0)

Description: X = multi-dimensional point data, y = labels (1 = outliers, 0 = inliers)