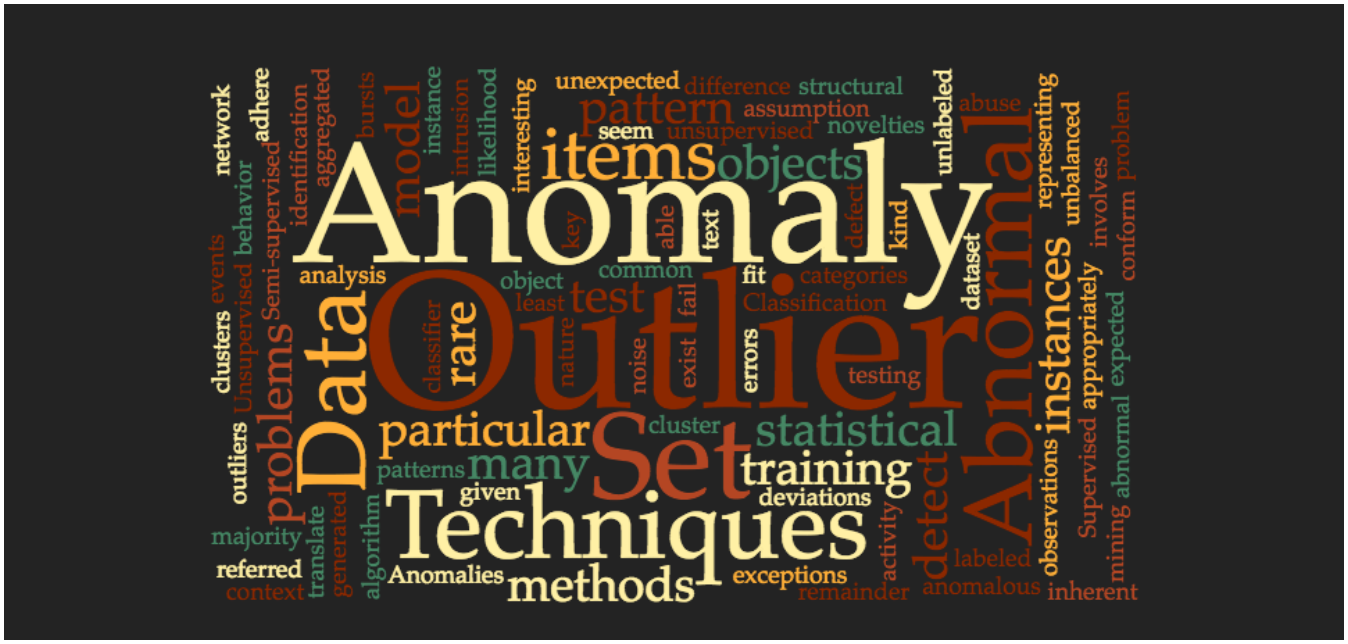

Thyroid Disease dataset

Dataset information

The original thyroid disease (ann-thyroid) dataset from UCI machine learning repository is a classification dataset, which is suited for training ANNs. It has 3772 training instances and 3428 testing instances. It has 15 categorical and 6 real attributes. The problem is to determine whether a patient referred to the clinic is hypothyroid. Therefore three classes are built: normal (not hypothyroid), hyperfunction and subnormal functioning. For outlier detection, 3772 training instances are used, with only 6 real attributes. The hyperfunction

class is treated as outlier class and other two classes are inliers, because hyperfunction is a clear minority class.

Source (citation)

F. Keller, E. Muller, K. Bohm."HiCS: High-contrast subspaces for density-based outlier ranking." ICDE, 2012.

C. C. Aggarwal and S. Sathe, "Theoretical foundations and algorithms for outlier ensembles." ACM SIGKDD Explorations Newsletter, vol. 17, no. 1, pp. 24–47, 2015.Downloads

Saket Sathe and Charu C. Aggarwal. LODES: Local Density meets Spectral Outlier Detection. SIAM Conference on Data Mining, 2016.

Downloads

File: thyroid.mat

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

Archives

Categories

- No categories