Visualization of LEGEND-200 data for high-dimensional classification algorithms

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

Miguel A. B. Schott, Esteban León, Julieta Gruszko, Aobo Li

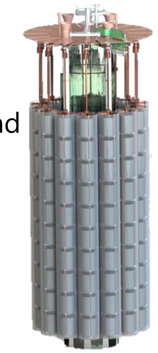
Department of Physics and Astronomy

LEGEND - 200

- The Large Enriched Germanium Experiment for Neutrinoless ββ (LEGEND) follows from previous neutrino experiments GERDA and MAJORANA.
- The goal is to detect the hypothetical nuclear decay

  (A,Z) → (A, Z+2) + 2e$^-$

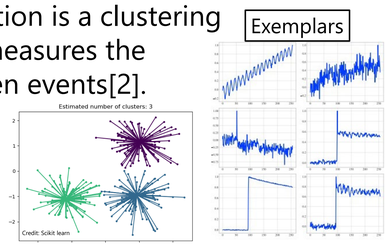
- LEGEND-200 overcomes backgrounds through a configuration of ^{76}Ge detectors with a total mass of 200kg.
- Observation of 0νββ would demonstrate Lepton Number Violation[1].



Credit: P. Krause

Affinity Propagation

- Affinity propagation is a clustering algorithm that measures the similarity between events[2].
- The center of a cluster is the exemplar.

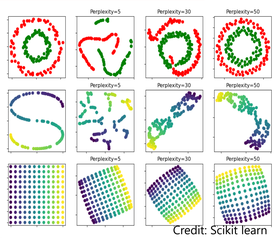


t – Distributed Stochastic Neighbor Embedding (t- SNE)

- t-SNE is used to visualize high-dimensional data[3].

Hyperparameters:

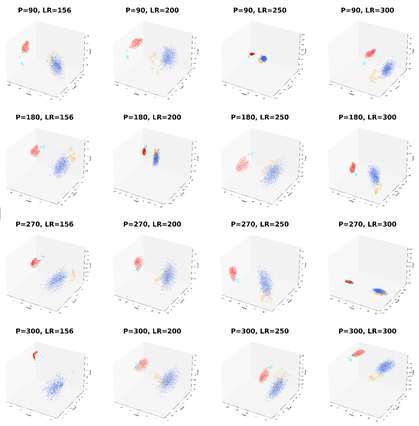
1. perplexity: number of neighbors considered for each point
2. learning rate: the step size of the algorithm

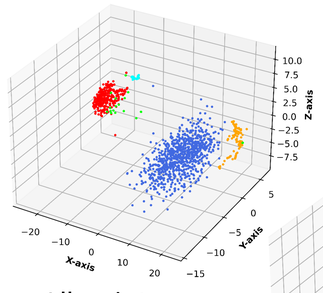


Perplexity may have dramatic impacts on the visualization

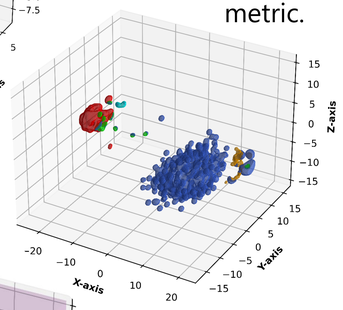
LEGEND – 60 Run 14 Data

L-60: Early Commissioning

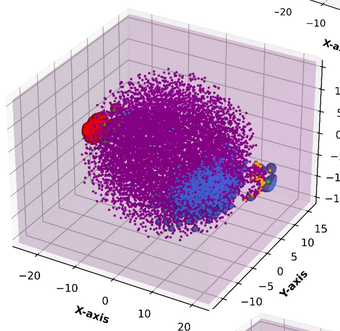




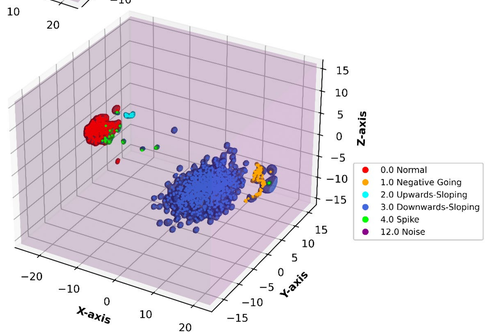
Non-noise points perplexity = 270, learning rate = 156, with the *Chebyshev* metric.



- All points are used to train a 3-D SVM



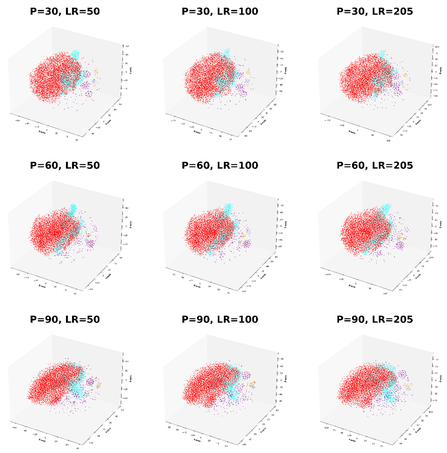
All points are shown. The noise points fall in the outer purple region.



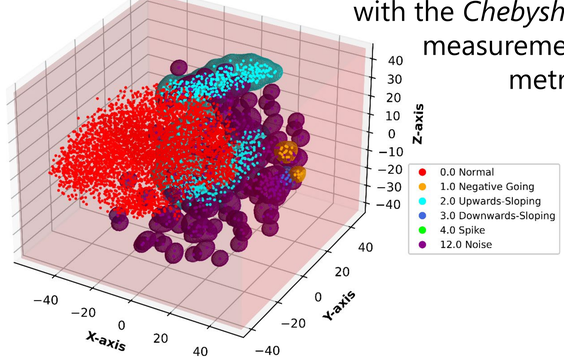
Noise points removed.

LEGEND – 200 Run 10 Data

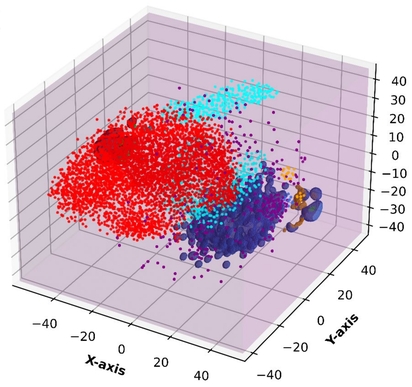
L-200: Full Experimental Commissioning



perplexity = 270, learning rate= 156, with the *Chebyshev* measurement metric.



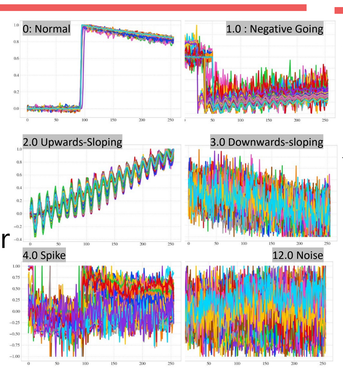
L-200 t-SNE overlapping L-60 SVM



Data Cleaning

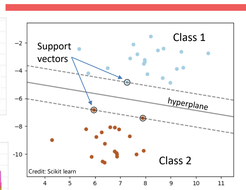
3-step traditional data cleaning process:

1. Identify undesirable data.
2. Create a parameter to isolate desirable results.
3. Cut data that falls outside the created parameters.



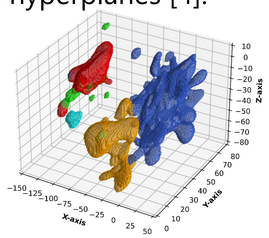
- Traditional data cleaning may unintentionally remove desirable waveforms
- The goal of the machine learning implementation is to reduce sacrifice while automating the classification of all physical events
- The machine learning approach uses Affinity Propagation (AP) and Support Vector Machine (SVM) algorithms.

Support Vector Machine (SVM)



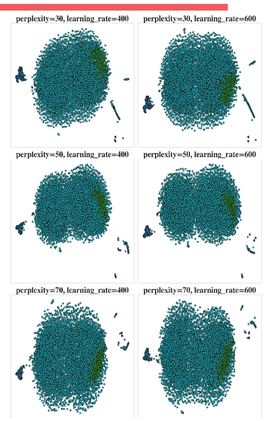
Support Vector Machines (SVM) classify objects by drawing hyperplanes [4].

The SVM for LEGEND-200 data was trained from the 3-D t-SNE algorithm.



Methodology

1. Different combinations of hyperparameters for t-SNE are made.
2. Hyperparameters are optimized
3. A 3-D SVM is trained on the selected t-SNE.
4. The SVM is rendered in voxels



References

[1] LEGEND Collaboration, LEGEND-200 First Stage, Online. https://legend-exp.org/science/legend-pathway/legend-200
[2] Frey BJ, Dueck D. Clustering by passing messages between data points. Science. 2007 Feb 16;315(5814):972-6. doi: 10.1126/science.1136800. Epub 2007 Jan 11. PMID: 17218491.
[3] van der Maaten, L.J.P.; Hinton, G.E. Visualizing High-Dimensional Data Using t-SNE. Journal of Machine Learning Research 9:2579-2605, 2008.
[4] "A Tutorial on Support Vector Regression" Alex J. Smola, Bernhard Schölkopf - Statistics and Computing archive Volume 14 Issue 3, August 2004, p. 199-222.