

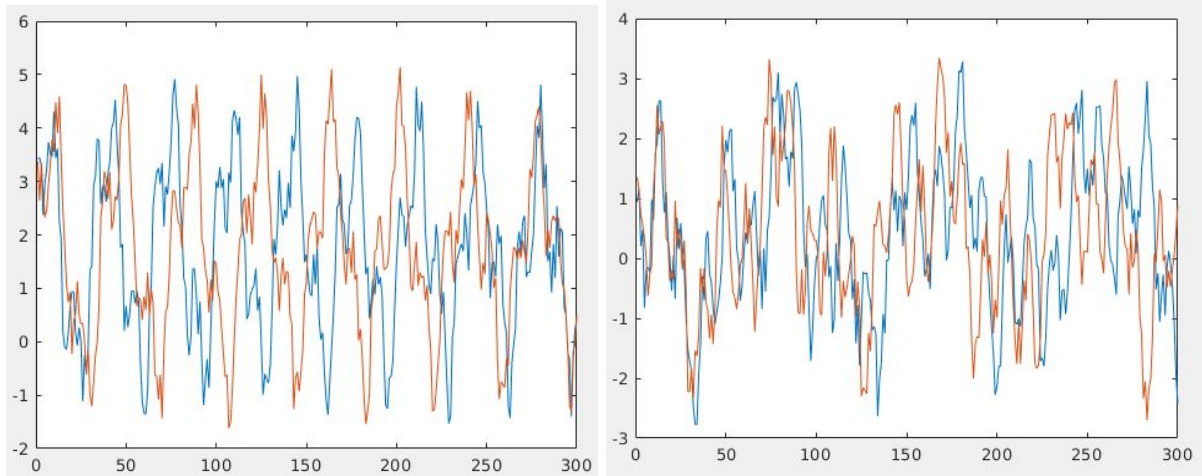
Algorithm:

Given:

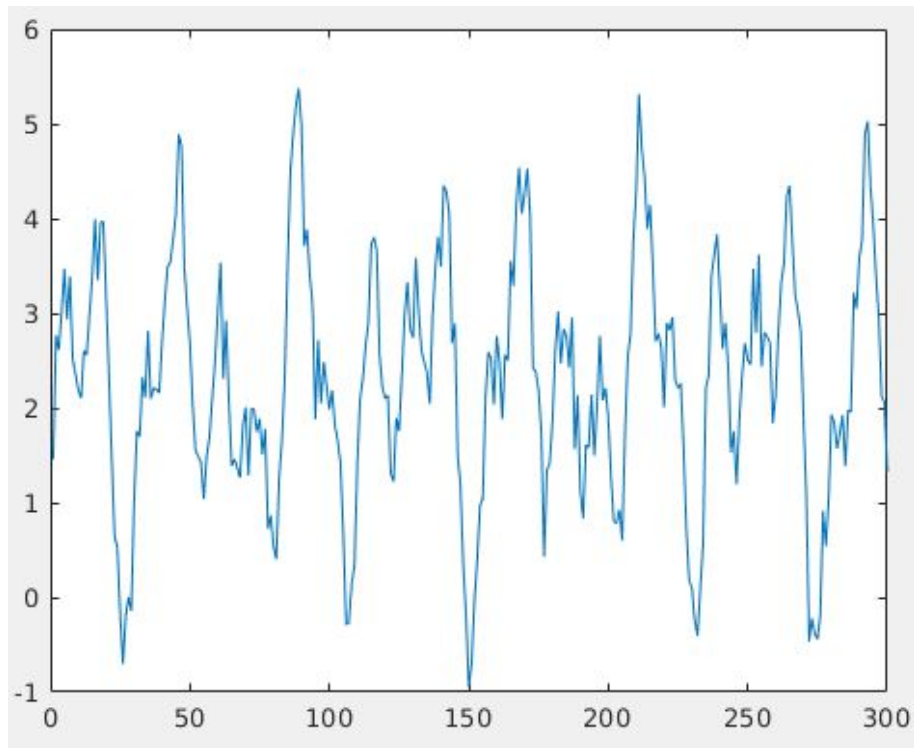
- S a set of data-points with n points of d_s dimensionality
- ϵ the minimum distance to be considered density reachable
- minPts the minimum ϵ -reachable neighbors to make a core point
- t , a threshold cluster size to divide noise from correct data
- d_t a target dimensionality for S

Do:

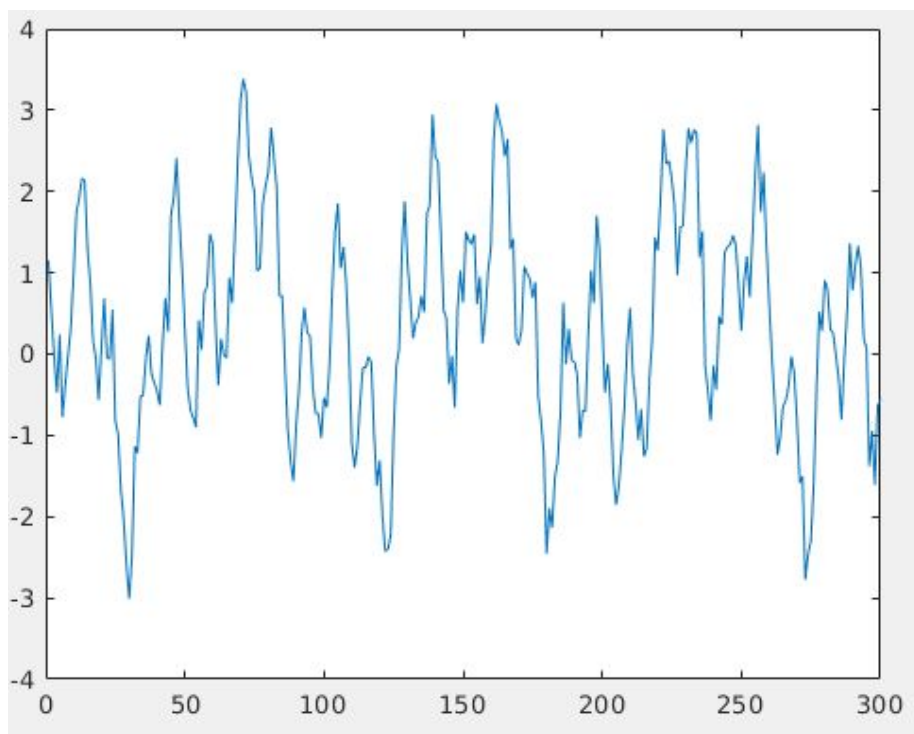
1. Train and deploy an autoencoder to generate set S' from S with n points of dimensionality d_t . Ignore this step is $d_s = d_t$.
2. For each s in S' append the label l which is observed for s to create S'' with n points of dimensionality d_t+1 .
3. Apply $\text{DBScan}(\epsilon, \text{minPts})$ to S'' to generate C , a set of clusters observed in S'' .
4. Add all points in C_o to M , a set of mislabeled example indexes from S .
5. For all remaining c in C , add all indexes in c to M if $|c| < t$.
6. Return M .



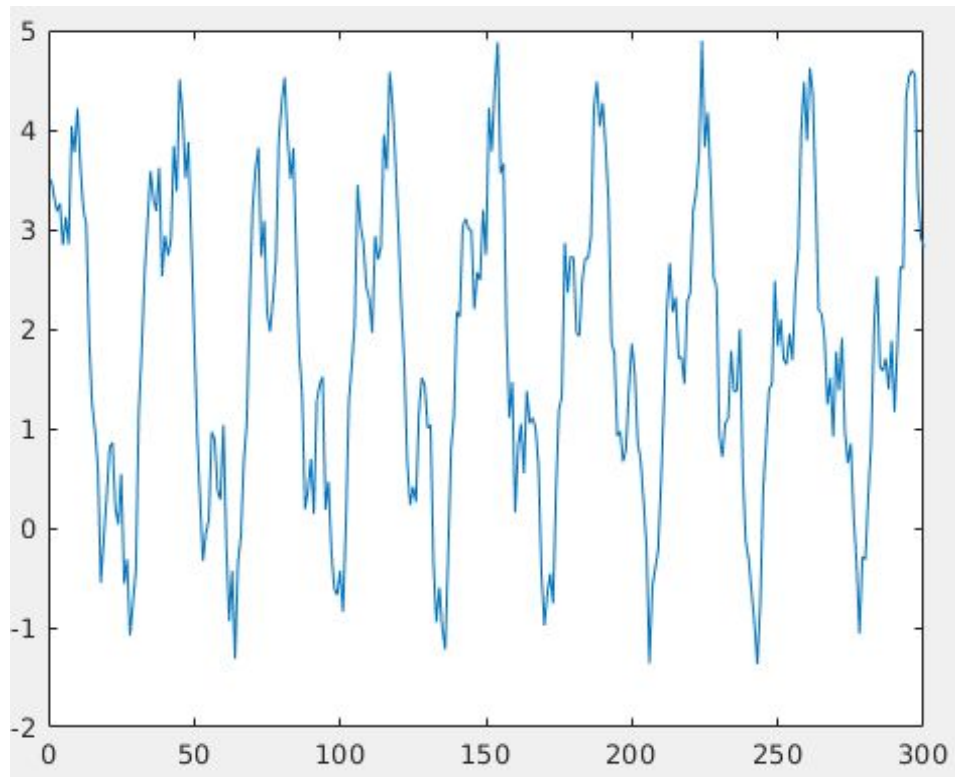
4 synthetic samples with same-labeled examples overlaid on each other.



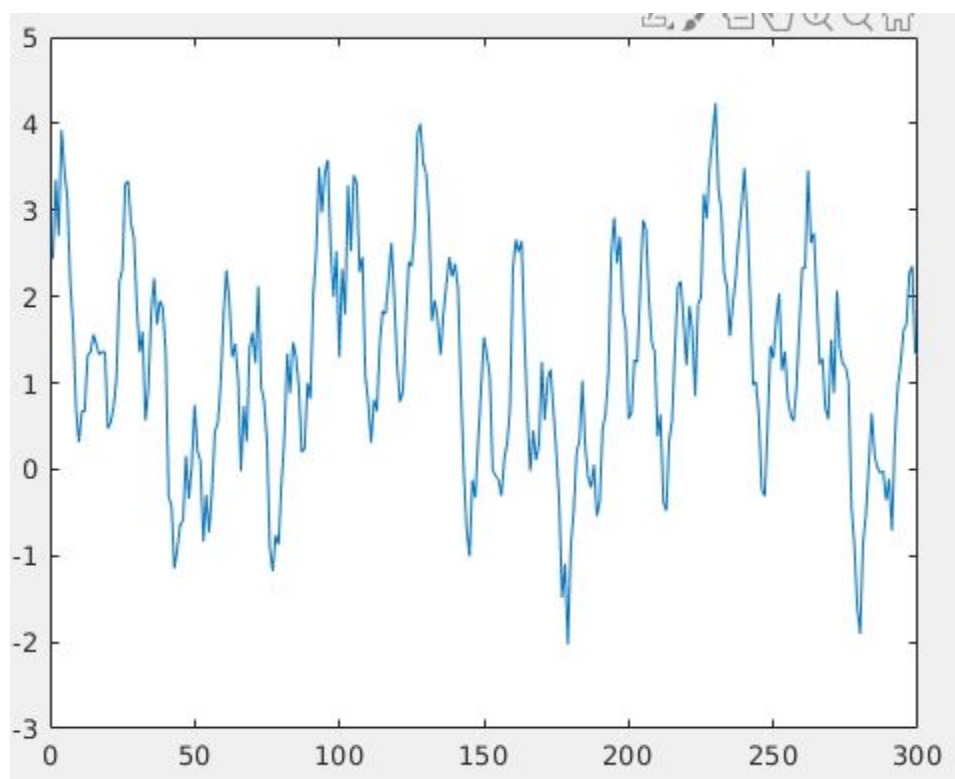
Label 0



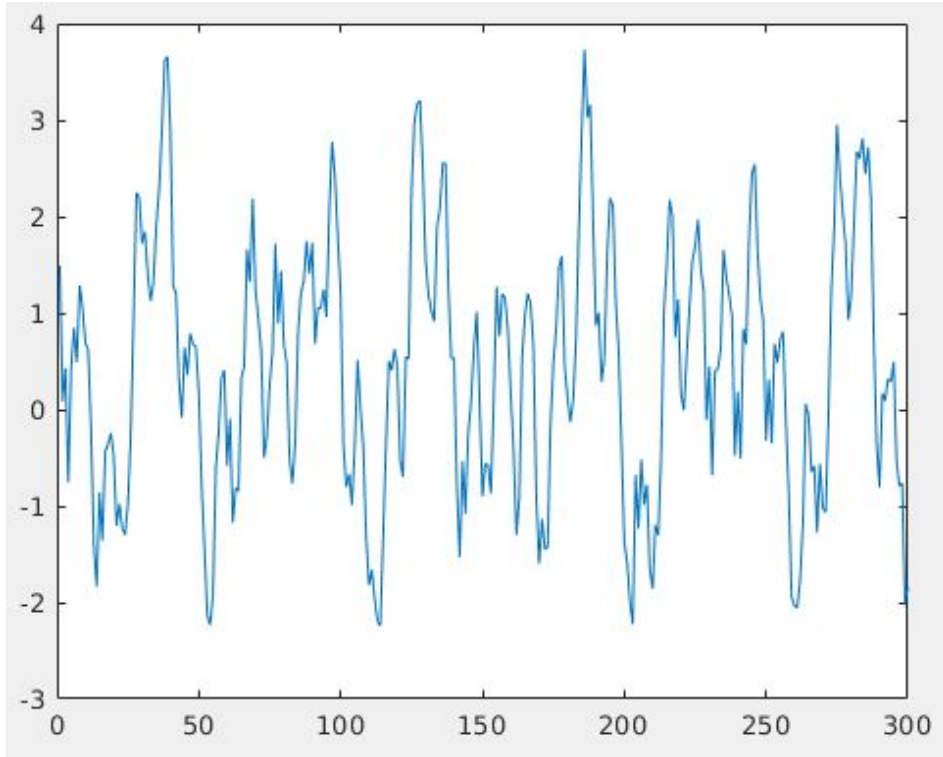
Label 1



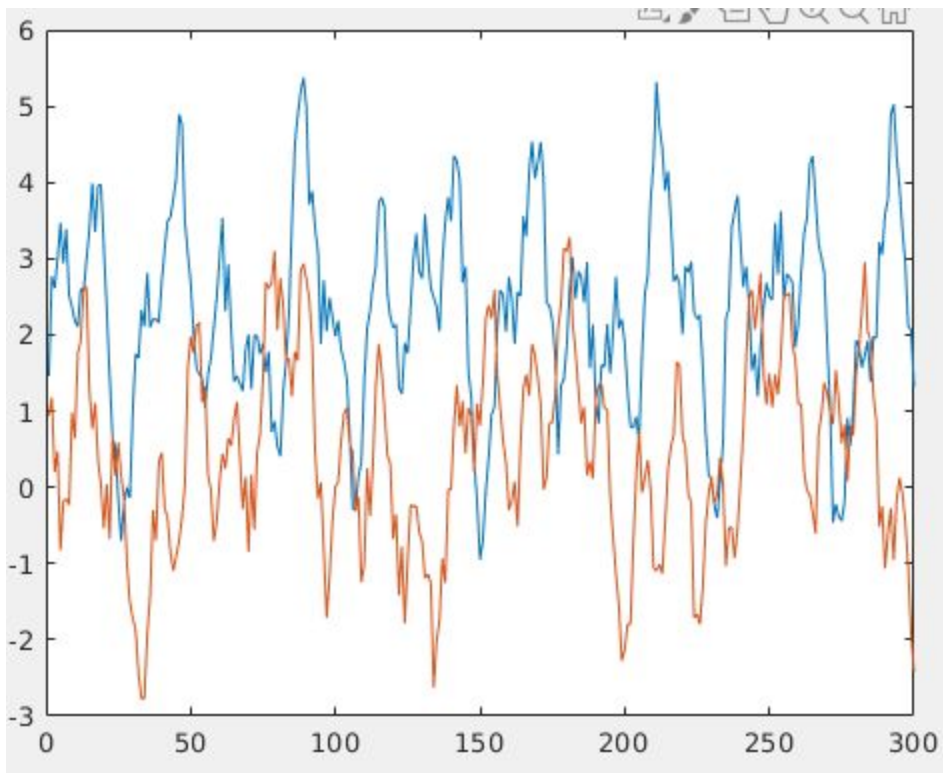
Label 2



Label 3



Label 4



A 0-labeled sample overlaid with a 1-labeled sample.