Input data:
$D = (D_L, D_U)$

Repeat N times

Phenotype modeling: Select D_L^i and model $F = D_L^i \cup D_U$ using a GMM (cluster number estimate using BIC).

Get candidats: Members of D_U clustered with D_L^i are considered candidats of class i.

next class i

Candidates classified as members of i are merged with D_L^i. Those which are rejected are returned to D_U

Each candidate is tested against an **outlier detection** algorithm.

all classes considered

Update classes: examples in D_U that are consistently accepted into a single class i are labelled as members of D_L^i.

New phenotype: Any example of D_U not merged with any D_L^i and which are consistenlty clustered together throughout the N iterations are considered members of a *new phenotype*.

Output: Return unassigned examples, new D_L^i members and new phenotypes.