Hannah Estes (hmestes), Saba Khanmohammadi (skhanmo), Mohsen EsfandyariDoulabi (mesfand)

```
➜  seai26spr git:(main) python3 -B my_hc.py 1 ~/gits/moot/optimize/misc/auto93.csv \
[  | sort -t, -k 3 -n | column -s, -t | cat -n
     1    100
     2    100
     3    15
     4    20
     5    25
     6    30
     7    35
     8    40
     9    45
    10    50
    11    55
    12    60
    13    65
    14    7
    15    70
    16    75
    17    80
    18    85
    19    90
    20    95
    21  :n    5 :lo  0.43 :mid  0.64 loop labelled
    22  :n    7 :lo  0.43 :mid  0.64 final labelled
    23  :n    7 :lo  0.43 :mid  0.64 loop labelled
    24  :n   10 :lo  0.43 :mid  0.64 loop labelled
    25  :n   15 :lo  0.43 :mid  0.64 final labelled
    26  :n   15 :lo  0.43 :mid  0.64 loop labelled
    27  :n   20 :lo  0.43 :mid  0.64 final labelled
    28  :n   20 :lo  0.43 :mid  0.64 loop labelled
    29  :n   25 :lo  0.43 :mid  0.64 final labelled
    30  :n   25 :lo  0.43 :mid  0.64 loop labelled
    31  :n   30 :lo  0.43 :mid  0.64 final labelled
    32  :n   30 :lo  0.43 :mid  0.64 loop labelled
    33  :n   35 :lo  0.43 :mid  0.64 final labelled
    34  :n   35 :lo  0.43 :mid  0.64 loop labelled
    35  :n   40 :lo  0.43 :mid  0.64 final labelled
    36  :n   40 :lo  0.43 :mid  0.64 loop labelled
    37  :n   45 :lo  0.43 :mid  0.53 final labelled
    38  :n   45 :lo  0.43 :mid  0.53 loop labelled
    39  :n   50 :lo  0.43 :mid  0.53 final labelled
    40  :n   50 :lo  0.43 :mid  0.53 loop labelled
    41  :n   55 :lo  0.43 :mid  0.53 final labelled
    42  :n   55 :lo  0.43 :mid  0.53 loop labelled
    43  :n   60 :lo  0.43 :mid  0.53 final labelled
    44  :n   60 :lo  0.43 :mid  0.53 loop labelled
    45  :n   65 :lo  0.43 :mid  0.53 final labelled
    46  :n   65 :lo  0.43 :mid  0.53 loop labelled
    47  :n   70 :lo  0.43 :mid  0.53 final labelled
    48  :n   70 :lo  0.43 :mid  0.53 loop labelled
    49  :n   75 :lo  0.43 :mid  0.53 final labelled
    50  :n   75 :lo  0.43 :mid  0.53 loop labelled
    51  :n   80 :lo  0.43 :mid  0.43 final labelled
    52  :n   80 :lo  0.43 :mid  0.43 loop labelled
    53  :n   85 :lo  0.43 :mid  0.43 final labelled
    54  :n   85 :lo  0.43 :mid  0.43 loop labelled
    55  :n   90 :lo  0.43 :mid  0.43 final labelled
    56  :n   90 :lo  0.43 :mid  0.43 loop labelled
    57  :n   95 :lo  0.43 :mid  0.43 final labelled
    58  :n   95 :lo  0.43 :mid  0.43 loop labelled
    59  :n  100 :lo  0.43 :mid  0.43 final labelled
    60  :n  100 :lo  0.43 :mid  0.43 final labelled
    61  :n  100 :lo  0.43 :mid  0.43 loop labelled
    62  :n  398 :lo  0.07 :mid  0.54 total rows
    63  0.09
    64  Convergence threshold: 0.09
```

Code

```
#!/usr/bin/env python3 -B
# ./rand.py $RANDOM ~/gits/moot/optimize/misc/auto93.csv

import random, sys, xai
xai.the.data=sys.argv[2]
random.seed(int(sys.argv[1]))
data = xai.Data(xai.csv(xai.the.data))
def Y(r): return round(xai.disty(data,r),2)
def top(a): a.sort(); return a[0]
def mid(a): a.sort(); n=len(a)//10; return a[5*n]
def sd(a):  a.sort(); n=len(a)//10; return (a[9*n]-a[n])/2.56
cohen = 0.35

def shuffle(rows):
    random.shuffle(rows)
    return rows

def report(what,rows):
    a=sorted(rows[:],key=Y)
    print(f":n {len(a):4} :lo {Y(a[0]):5.2f} :mid {Y(a[len(a)//2]):5.2f}", what)

def report_lowest(what, rows):
    a=sorted(rows[:],key=Y)
    print(f":lo {Y(a[0]):5.2f}", what)

def extremes(rows):
    a=sorted(rows[:],key=Y)
    n=len(rows)//10
    ok = a[n]
    no = a[1-n]
    return ok, no

def project(r, ok, no):
  c=xai.distx(data,ok,no)
  a=xai.distx(data,r,ok)
  b=xai.distx(data,r,no)
  return (a**2 + c**2 - b**2) / (2*c + 1e-32)

def prune(rows, ok, no):
  a=sorted(rows[:],key=lambda r: project(r, ok, no))
  return a[:len(a)//2]

fn = sd  # or mid or top
eps = fn([Y(r) for r in data.rows]) * cohen
print(f"Convergence threshold: {eps:.2f}")

budget = 100
step = 5
labelled, rows = [], shuffle(data.rows[:])
b4 = 1e32

report("total rows", rows)
print(round(eps, 2))
for _ in range(20):
  while len(labelled) < budget:
    labelled += shuffle(rows)[:step]
    ok, no = extremes(labelled)
    rows = prune(labelled, ok, no)
    report("loop labelled", labelled)
    now = fn([Y(r) for r in labelled])
    if abs(b4 - now) > eps:
      b4 = now
    else:
      break
  print(", ", len(labelled), end=", \n")
  report("final labelled", labelled)
  print("")
```