```lua
-- vim: filetype=lua ts=2 sw=2 et:
-- (c) 2022, Tim Menzies,  timm@ieee.org, opensource.org/licenses/Fair
-- Usage of the works is permitted provided that this instrument is retained
-- with the works, so that any entity that uses the works is notified of this
-- instrument.  DISCLAIMER: THE WORKS ARE WITHOUT WARRANTY.
--     __                                        __
--    /\ \__                                    /\ \
--    \ \ ,_\    __  __  __     __      _       \ \ \/'\
--     \ \ \/  /\ \/\ \/\ \  /'__`\  /'__`\     \ \ , <
--      \ \ \_ \ \ \_/ \_/ \/\  __/ /\ \L\.\_   \ \ \\`\
--       \ \__\ \ \___x___/'\ \____\\ \__/.\_\   \ \_\ \_\
--        \/__/  \/__//__/   \/____/ \/__/\/_/    \/_/\/_/
local help= [[
tweak: tries three weak learners for multi-objective optimization
(c) 2022, Tim Menzies,  timm@ieee.org, opensource.org/licenses/Fair

learner1: n times, discard half the data furthest from best
learner2: classify data according to presence of survivors of learner1
learner3: run learner1 on best "best" found by learner2

USAGE:
  alias twk="lua tweak.lua "
  twk [OPTIONS]

OPTIONS:
  --cohen   -c  cohen                      = .35
  --K       -K  manage low class counts    = 1
  --M       -M  manage low evidence counts = 2
  --far     -F  how far to go for far      = .9
  --p       -p  coefficient on distance    = 2
  --seed    -S  seed                       = 10019
  --some    -s  sample size for distances  = 512
  --stop    -T  how far to go for far      = 20
  --min     -m  size of min space          = .5
  --best    -B  best percent               = .05

OPTIONS (other):
  --dump    -d  dump stack+exit on error   = false
  --file    -f  file name                  = ../etc/data/auto93.csv
  --help    -h  show help                  = false
  --rnd     -r  rounding numbers           = %5.3f
  --go      -g  start up action            = nothing]] --[[

ABOUT THE CODE:
- Settings generated from "help" string
  - Settings can be updated from the strings seed in flags
  - Settings stored in the global "the"

- Layout code in chunks of size 120 lines (max), broken by line-feed
  - Chunk1=header; Chunk2=utils; Chunk3=objects; Chunk(last)=demos+start-up
- Layout lines 80 chars side (max)
  - So use 2 spaces for "tab"
- Do functions as one-liners (if possible)
- In order to define code in any order:
  - Near the top, define all function and Object names as "local"
- Otherwise, don't use the "local" keyword (too ugly)

- Minimize use of map (hard to debug)
- Object names are short and UPPER CASE
- Private object slots (that should not be printed) start with "_".
- Constructors need not return constructed instance.
- No inheritance (hard to debug)
- For code with many parameters, pass in a dictionary with named fields.

- Tests  in the "go" table at end. Reset settings to defaults after each.
- Tests check for error conditions using "ok" not "assert".
- Command line "-d -go x" crashed if test "x" fails, shows stack dump.
- Command line "-go x" calls test "go.x()"
- Command line "-go all" calls all tests.
- Command line "-h" shows help
- Command line "-S N" sets random seed (so "-S $RANDOM" is "full" random)

- 2nd last line: look for "rogue" globals (there should be none)
- Last line: exit to operating system with number of failures seen in tests

ABOUT THE LEARNERS:
- "mean","mode" are generalized to "mid" (i.e. "mid-point")
- "standard deviation","entropy" are generalized  to "div" (i.e. "diversity")

- BIN holds the class labels seen between "lo" and "hi".
- EGS (examples) hold many ROWs, summarized in SYMbolic or NUMeric  columns.
- COLS is a factory for turning column names into NUMs or SYMs.
  - Numeric names start with upper case
  - Goal names ending with "-" or ""+ get weights -1,1 for minimize,maximize
  - Non-numeric class names end with "!"
  - Columns to be skipped have a name ending with ":"
  - Non-skipped columns are divided into COLS.y and COLS.x (for goal and other)
- "mid" and "div" for EGS are computed recursively by "mid,div" in NUMs,SYMs
- distances between two rows is computed recursively via "dist" in NUMs,SYMs

- ROW1 before ROW2 (i.e. ROW1<ROW2) if its goals dominate (using [CDOM])
- ROWs are recursively separated and clustered by [FASTMAP] random projections
- The distance between two ROWs (i.e. ROW1-ROW2) uses [AHA].
- ROWs are persistent. They are created once but many be used in many EGS.
- ROWs have a "_data" pointer where it gets "lo,hi" info needed for distances.
- For consistency's sake, ROW._data is fixed to the first
  EGS that holds that row.

REFERENCES:
- [AHA]    : Aha       : doi.org/10.1007/BF00153759
- [CDOM]   : Zitzler   : doi.org/10.1145/1830483.1830578
- [FASTMAP]: Faloutsos : doi.org/10.1145/568271.223812      --]]
```

```lua
---     _   _  _____  _  _      ____
---    | | | ||_   _|| || |    [__
---    |_| |   | |  | || |___  ___]

local the,any,cells,csv,fmt,fu,lt,many,map = {}
local oo,o,obj,per,push,R,rnd,rnds,sort,slice,string2thing
---                                           _
---     _  -|-  ._  o  ._    _    _     )    -|-  |_   o  ._    _    _
---    _>   |_  |   |  | |  (_|  _>    /_     |_  | |  |  | |  (_|  _>
---                          _|                                  _|
function string2thing(x)
  x = x:match"^%s*(.-)%s*$"
  if x=="true" then return true elseif x=="false" then return false end
  return math.tointeger(x) or tonumber(x) or x end

help:gsub("\n ([-][-]([^%s]+))[%s]+(-[^%s]+)[^\n]*%s([^%s]+)",function(f1,k,f2,x)
  for n,flag in ipairs(arg) do if flag==f1 or flag==f2 then
    x = x=="false" and"true" or x=="true" and"false" or arg[n+1] end end
  the[k] = string2thing(x) end)

function csv(src)
  src = io.input(src)
  return function(line, row)
    line=io.read()
    if not line then io.close(src) else
      row={}; for x in line:gmatch("([^,]+)") do row[1+#row]=string2thing(x) end
      return row end end end

---     ._ _   o   _   _
---     | | |  |  _>  (_
R = math.random
function fu(x) return function(a)   return a[x] end end
function lt(x) return function(a,b) return a[x] < b[x] end end

---     |  o   _  -|-   _
---     |  |  _>   |_  _>
function any(a, i)    i=R()*#a//1; i=math.max(1,math.min(i,#a)); return a[i] end
function many(a,n, u) u={}; for j=1,n do        u[1+#u]= any(a) end;return u end
function map(t,f, u)  u={}; for _,v in pairs(t) do u[1+#u]=f(v) end;return u end
function per(t,p)     return t[ ((p or .5)*#t) // 1 ] end
function push(t,x)    t[1+#t]=x; return x end
function sort(t,f)    table.sort(t,f) return t end
function slice(t,i,j,   u)
  u={}; for k=(i or 1), (j or #t) do u[1+#u] = t[k] end return u end

---     ._   ._  o  ._   -|-
---     |_)  |   |  | |   |_
---     |
fmt = string.format
function oo(t) print(o(t)) end
function o(t,    u,one,hide,sorted)
  if type(t) ~= "table" then return tostring(t) end
  sorted = #t>0 -- true when array's indexes are 1,2...#t
  hide= function(k) return tostring(k):sub(1,1) == "_" end
  one = function(k,v) return sorted and tostring(v) or fmt(":%s %s",k,v) end
  u={}; for k,v in pairs(t) do if not hide(k) then u[1+#u] = one(k,v) end end
  return (t.is or "").."{"..table.concat(sorted and u or sort(u)," ").."}" end

function rnds(t,f) return map(t, function(x) return rnd(x,f) end) end
function rnd(x,f)
  return fmt(type(x)=="number" and (x~=x//1 and f or the.rnd) or"%s",x) end

---     ._    _   |         ._ _    _   ._  ._   |_   o   _  ._ _
---     |_)  (_)  |  \/     | | |  (_)  |   |_)  | |  |  _>  | | |
---     |            /                      |
function obj(name,    t,new)
  function new(kl,...) local x=setmetatable({},kl); kl.new(x,...); return x end
  t = {__tostring=o, is=name or ""}; t.__index=t
  return setmetatable(t, {__call=new}) end
```

```lua
---     ____  ___    _  ____  ____  ___  ____
---    |    ||__]   | |___  |      |  [__
---    |____||__]  _| |___  |___   |  ___]

local SYM,BIN,NUM,COLS = obj"SYM",obj"BIN",obj"NUM",obj"COLS"
local ROW,EGS = obj"ROW",obj"EGS"

---      _        ._ _
---     _>  \/   | | |
---         /
function SYM:new(pos,s)
  self.pos, self.txt= pos or 0,s or ""
  self.n, self.has, self.most, self.mode = 0,{},0,nil end

function SYM:sub(x,inc) return self:add(x, -(inc or 1)) end
function SYM:add(x,inc)
  if x ~= "?" then
    inc = inc or 1
    self.n = self.n + inc
    self.has[x] = (self.has[x] or 0) + inc
    if self.has[x] > self.most then self.most,self.mode = self.has[x],x end end
  return x end

function SYM:mid() return self.mode end
function SYM:div(   e)
  e=0; for _,m in pairs(self.has) do
         if m>0 then e = e-m/self.n * math.log(m/self.n,2) end end
  return e end

function SYM:dist(x,y) return x=="?" and y=="?" and 1 or x==y and 0 or 1 end

function SYM:bins(rows,     out,known,x)
  out,known = {},{}
  for _,row in pairs(rows) do
    x = row.cells[self.pos]
    if x~="?" then
      known[x] = known[x] or push(out, BIN({txt=self.txt,  pos=self.pos,
                                            lo=x ,hi=x, ys=SYM()}))
      known[x].ys:add(row.klass) end end
  return out end

---     ._        ._ _
---     | |  |_|  | | |
function NUM:new(pos,s)
  self.pos, self.txt, self.lo, self.hi = pos or 0,s or "",1E32, -1E32
  self.n, self.mu, self.m2 = 0,0,0
  self.w = self.txt:find"-$" and -1 or 1  end

function NUM:add(x,   _,d)
  if x ~="?" then
    self.n  = self.n + 1
    self.lo = math.min(x, self.lo)
    self.hi = math.max(x, self.hi)
    d       = x - self.mu
    self.mu = self.mu + d/self.n
    self.m2 = self.m2 + d*(x - self.mu) end
  return x end

function NUM:mid() return self.mu end
function NUM:div() return (self.m2/(self.n - 1))^0.5 end

function NUM:norm(x,   lo,hi)
  lo,hi= self.lo, self.hi
  return x=="?" and x or hi-lo < 1E-9 and 0 or (x - lo)/(hi - lo) end

function NUM:dist(x,y)
  if     x=="?" and y=="?" then return 1 end
  if     x=="?"            then y = self:norm(y); x = y<.5 and 1 or 0
  elseif y=="?"            then x = self:norm(x); y = x<.5 and 1 or 0
  else x,y = self:norm(x), self:norm(y) end
  return math.abs(x - y) end

function NUM:bins(rows,      xy,div,xys,epsilon,small,b4,out)
  function xy(row, x)
    x=row.cells[self.pos]; if x~="?" then return {x=x,y=row.klass} end end
  function div(lo,hi,         x,y,cut,lhs,rhs,tmp,best,overall)
    lhs, rhs, overall = SYM(), SYM(), SYM()
    for i=lo,hi do overall:add( rhs:add(xys[i].y) ) end
    best = rhs:div()
    for i=lo,hi do
      x, y = xys[i].x, xys[i].y
      lhs:add( rhs:sub( y) )
      if lhs.n > small and rhs.n > small then
        if x ~= xys[i+1].x then
          if x - xys[lo].x > epsilon and xys[hi].x - x > epsilon then
            tmp = (lhs.n*lhs:div() + rhs.n*rhs:div())  / (lhs.n + rhs.n)
            if tmp < best then
              best,cut = tmp,i end end end end end
    if   cut
    then div(lo,    cut)
         div(cut+1, hi)
    else b4= push(out, BIN({txt=self.txt, pos=self.pos, lo=b4,
                            hi=xys[hi].x, ys=overall})).hi end
  end ----------------------------
  xys     = sort(map(rows,xy), lt"x")
  b4,out  = -math.huge, {}
  epsilon = (per(xys,.9).x - per(xys,.1).x) / 2.56*the.cohen
  small   = (#xys)^the.min
  div(1, #xys)
  out[#out].hi = math.huge
  return out end
```

```lua
---
---     _   _   |   _
---    (_  (_)  |  _>

function COLS:new(names,       it,num,sym,col)
  self.names, self.x, self.y, self.all = names, {},{},{}
  for pos,txt in pairs(names) do
    col = push(self.all, (txt:find"^[A-Z]" and NUM or SYM)(pos,txt))
    if not txt:find":$" then
      if txt:find"!$" then self.klass = col end
      push(txt:find"[-+!]$" and self.y or self.x, col) end end end

---    |_   o  ._
---    |_)  |  | |
function BIN:new(t)
  self.pos, self.txt  = t.pos, t.txt
  self.lo, self.hi, self.ys = t.lo, t.hi, t.ys or SYM() end

function BIN:__tostring()
  local x,lo,hi,big = self.txt, self.lo, self.hi, math.huge
  if     lo ==  hi  then return fmt("%s == %s",x, lo)
  elseif hi ==  big then return fmt("%s >= %s",x, lo)
  elseif lo == -big then return fmt("%s < %s", x, hi)
  else                   return fmt("%s <= %s < %s",lo,x,hi) end end

function BIN:select(t)
  t = t.cells and t.cells or t
  local x, lo, hi = t[self.pos], self.lo, self.hi
  return x=="?" or lo == hi and lo == x or lo <= x and x < hi end

---    ._   _
---    |   (_)  \/\/
function ROW:new(data,t)
  self._data,self.cells, self.evaluated = data,t, false end

function ROW:__sub(other,    cols,d,inc)
  d, cols = 0, self._data.cols.x
  for _,col in pairs(cols) do
    inc = col:dist(self.cells[col.pos], other.cells[col.pos])
    d   = d + inc^the.p end
  return (d / #cols) ^ (1/the.p) end

function ROW:__lt(other,   s1,s2,e,y,a,b)
  y= self._data.cols.y
  s1, s2, e = 0, 0,  math.exp(1)
  for _,col in pairs(y) do
     a = col:norm(self.cells[col.pos])
     b = col:norm(other.cells[col.pos])
     s1= s1 - e^(col.w * (a - b) / #y)
     s2= s2 - e^(col.w * (b - a) / #y) end
  return s1/#y < s2/#y  end

---     _    _   _
---    (/_  (_|  _>
---          _|
function EGS:new() self.rows,self.cols = {},nil end

function EGS:add(t)
  if   self.cols
  then t = push(self.rows, t.cells and t or ROW(self,t)).cells
       for _,col in pairs(self.cols.all) do col:add(t[col.pos]) end
  else self.cols = COLS(t) end
  return self end

function EGS:mid(t) return map(t or self.cols.y,function(c)return c:mid()end)end
function EGS:div(t) return map(t or self.cols.y,function(c)return c:div()end)end

function EGS:clone(rows, out)
  out=EGS():add(self.cols.names)
  for _,row in pairs(rows or {}) do out:add(row) end
  return out end

function EGS:load(file)
  for t in csv(file) do self:add(t) end
  return self end

function EGS:around(r1,rows,   t)
  t={}; for _,r2 in pairs(rows or self.rows) do push(t,{row=r2, d= r1 - r2}) end
  return sort(t,lt"d") end

function EGS:far(r1,rows)
  return per(self:around(r1,rows),the.far).row end

function EGS:sway(rows,stop,rest,x,           some,y,c,best,mid)
  rows = rows or self.rows
  stop = stop or 2*the.best*#rows
  rest = rest or {}
  if #rows <= stop then return rows,rest end
  some = many(rows,the.some)
  x    = x or self:far(any(some), some)
  y    =      self:far(x,         some)
  if y < x then x,y = y,x end -- "x" is now better than "y"
  x.evaluated = true
  y.evaluated = true
  c    = x - y
  rows = map(rows,function(r) return {r=r, x=((r-x)^2+c^2-(r-y)^2)/(2*c)} end)
  best = {}
  mid  = #rows//2
  for i,rx in pairs(sort(rows,lt"x")) do push(i<=mid and best or rest, rx.r) end
  return self:sway(best,stop,rest,x)  end
---
---   ___  ____ _  _ ____ ____
---   |  \ |___ |\/| |  | [__
---   |__/ |___ |  | |__| ___]

local go,no,fails,ok,main={},{},0

function main(   all,b4)
  all={}; for k,_ in pairs(go) do push(all,k) end
  for _,x in pairs(the.go=="all" and sort(all) or {the.go}) do
    b4={}; for k,v in pairs(the) do b4[k]=v end
    math.randomseed(the.seed)
    if go[x] then print(x); go[x]() end
    for k,v in pairs(b4) do the[k]=v end end end

function ok(test,msg)
  print("", test and "PASS "or "FAIL ", msg or "")
    if not test then
      fails= fails+1
      if the.dump then assert(test,msg) end end end

function go.rogue( t)
  t={}; for _,k in pairs{ "_G", "_VERSION", "arg", "assert", "collectgarbage",
  "coroutine", "debug", "dofile", "error", "getmetatable", "io", "ipairs",
  "load", "loadfile", "math", "next", "os", "package", "pairs", "pcall",
  "print", "rawequal", "rawget", "rawlen", "rawset", "require", "select",
  "setmetatable", "string", "table", "tonumber", "tostring", "type", "utf8",
  "warn", "xpcall"} do t[k]=true end
  for k,v in pairs(_ENV) do if not t[k] then print("?",k, type(v)) end end end

function go.the()  oo(the) end
function go.eg( n,out)
  out =true
  n=0; for row in csv(the.file) do
         n=n+1; out=out and #row==8
                if n>1 then out=out and type(row[1])=="number" end end
  ok(out and n==399); end

function go.rows( egs)
  egs=EGS():load(the.file)
  map(egs.cols.x,oo); print("");
  map(egs.cols.y,oo) end

function go.dist( egs,    a,b,c,out)
  egs   = EGS():load(the.file)
  out = true
  for i=1,100 do
    a,b,c = any(egs.rows), any(egs.rows), any(egs.rows)
    out   = out and (b -a)==(a-b) and (a-a)==0 and ((a-b)+ (b-c) >= (a-c)) end
  ok(out,"dist") end

function go.sort(    egs,rows,n)
  egs   = sort(EGS():load(the.file))
  rows= sort(egs.rows)
  n = .05*#rows//1
  print("what", o(map(egs.cols.y,function(c) return c.txt end)))
  print("all",   o(rnds(egs:mid())))
  print("best", o(rnds(egs:clone(slice(rows, 1, n)):mid())))
  print("rest", o(rnds(egs:clone(slice(rows, n+1 )):mid())))
  end

function go.far(  egs,row2)
  egs = EGS():load(the.file)
  row2=egs:far(egs.rows[1])
  print(row2 - egs.rows[1])end

function go.sway(  egs,best,rest)
  egs = EGS():load(the.file)
  best,rest = egs:sway()
  for _,row in pairs(egs.rows) do if row.evaluated then oo(row.cells) end end
  print("all",   o(rnds(egs:mid())))
  print("best", o(rnds(egs:clone(best):mid())))
  print("rest", o(rnds(egs:clone(rest):mid()))) end

function go.symbins( egs)
  egs = EGS():load(the.file)
  for i,row in pairs(sort(egs.rows)) do row.klass = i<=#egs.rows//2 end
  map(egs.cols.x[4]:bins(egs.rows),oo) end

function go.numbins( egs)
  egs = EGS():load(the.file)
  for i,row in pairs(sort(egs.rows)) do row.klass = i<=.05*#egs.rows end
  for _,col in pairs(egs.cols.x) do
    print(fmt("\n%s",col.txt))
    map(col:bins(egs.rows),oo) end end

function go.classify(  egs,best,rest)
  egs = EGS():load(the.file)
  oo(egs.cols.x[4])
  best,rest = egs:sway()
  for _,row in pairs(egs.rows) do if row.evaluated then oo(row.cells) end end
  print("all",   o(rnds(egs:mid())))
  print("best1", o(rnds(egs:clone(best):mid())))
  print("rest1", o(rnds(egs:clone(rest):mid())))
  for _,row in pairs(best) do row.klass=true end end
  --for _,row in pairs(many(rest, 3*#best)) do

 ----------------------------------------------------------------------------
---
---   ____ ___ ____ ____ ___
---   [__   |  |__| |__/  |
---   ___]  |  |  | |  \  |

if the.help then print(help) else main() end
go.rogue()
os.exit(fails)
```