

```

1 local help=[
2 LOOK: landscape analysis
3 (c) 2022 Tim Menzies, tim@ieee.org, BSD2 license
4 "I think the highest and lowest points are the important ones.
5 Anything else is just... in between." ~Jim Morrison
6
7 INSTALL: requires: lua 5.4+
8 download: lib.lua, look.lua, looking.lua
9 test : lua egs.lua -h
10
11 USAGE: lua looking.lua [OPTIONS]
12
13                                     defaults
14 --Far -F far                        = .95
15 --How -H how we optimize: more,less,tabu = more
16 --Min -M min size pass2            = 10
17 --Some -S sample size              = 512
18 --also -a size of rest=best*also   = 4
19 --min -m min size pass1            = .5
20 --p -p distance coefficient        = 2
21 --seed -s random number seed      = 10019
22
23 --file -f csv file with data       = ../etc/data/auto93.csv
24 --go -g start up action            = nothing
25 --help -h show help               = false
26 --loud -l verbose mode            = false]]
27
28 local _ = require"lib"
29 local any,big,csv,is,lt,many,map=_.any,_.big,_.csv,_.is,_.lt,_.many,_.map
30 local o,oop,per,push,shuffle,sort=_.o,_.oo,_.per,_.push,_.shuffle,_.sort
31 local tothing =_.tothing
32
33 local the={}
34 help:gsub("[^-][^%s+](^n)%(^%s+)",function(k,x) the[k]=_.tothing(x)end)
35
36

```

```

36 local function nump(s) return s:find"^[A-Z].*" end
37 local function skip(s) return s:find"$" end
38 local function goalp(s) return s:find"[a-z$]" end
39 local function wght(s) return s:find"$" and ~1 or 1 end
40 local function ranges(col,...)
41 local ranges,tmp={},{}
42 for klass,row in pairs(rows) do
43 for _,row in pairs(rows) do
44 local v=row[col.at]
45 if v=="?" then
46 local bin = col:bin(v)
47 tmp[bin] = tmp[bin] or push(ranges, RANGE(x,x,SYM(col.at, col.txt)))
48 tmp[bin]:add(v,klass) end end end
49 return col:binsMerge(sort(ranges, lt"lo"))
50
51 -----
52 local RANGE,ROWS,TREE = is"RANGE", is"ROWS", is"TREE"
53 local ROW,SYM,NUM = is"ROW", is"SYM", is"NUM"
54
55 function ROW.new(i,of,cells) i,cells, i.of, i.evaluated = cells,of,false end
56 function ROW.__lt(i,j)
57 s1,s2, n = 0, 0, #i.of.y
58 for _,col in pairs(i.of.y) do
59 v1,v2 = col:norm(i.cells[col.at]), col:norm(j.cells[col.at])
60 s1 = s1 - 2.7183*(col.w * (v1 - v2) / n)
61 s2 = s2 - 2.7183*(col.w * (v2 - v1) / n) end
62 return s1/n < s2/n end
63
64 function ROW.dist(i,j, d,n)
65 d,n = 0,0
66 for _,col in pairs(i.of.xs) do
67 n = n+1
68 d = d + (col:dist(i.cells[col.at], j.cells[col.at]))^the.p end
69 return (d/d^n)^(1-the.p)
70
71 function SYM.new(i,at,txt)
72 i.at=at or 0; i.txt=txt or ""; i.all, i.n, i.most, i.mode = {},0,0,nil end
73
74 function SYM.dist(i,v1,v2)
75 return (v1=="?" and v2=="?" and 1) or (v1==v2 and 0 or 1) end
76
77 function SYM.add(i,v,n)
78 n = n or 1
79 if v == "?" then i.n=i.n+1; i.all[v] = n + (i.all[v] or 0);
80 if i.all[v]>i.most then i.most,i.mode = i.all[v],v end end end
81
82 function SYM.div(i, e)
83 e=0; for k,n in pairs(i.all) do e=e/n/i.n*math.log(n/i.n,2) end ;return e end
84
85 function SYM.mid(i) return i.mode end
86 function SYM.bin(i,x) return x end
87 function SYM.binsMerge(i,ranges) return ranges end
88
89 function SYM.merged(i,j,min, k)
90 k = SYM(i.at,i.txt)
91 for v,n in pairs(i.all) do k:add(v,n) end
92 for v,n in pairs(j.all) do k:add(v,n) end
93 min = min or 0
94 if i.n < min or j.n < min or k:div() < (i.n*i:div() + j.n*j:div())/k.n then
95 return k end end
96
97 function RANGE.new(i,lo,hi,y) i.lo,i.hi,i.y = lo, hi, y end
98
99 function RANGE.__tostring(i)
100 local x, lo, hi = i.y.txt, i.x.lo, i.x.hi
101 if lo == hi then return fmt("%s==%s",x, lo)
102 elseif hi == big then return fmt("%s>=%s",x, lo)
103 elseif lo == -big then return fmt("%s<=%s",x, hi)
104 else return fmt("%s<=%s<%s",lo,x,hi) end end
105
106 function RANGE.add(i,v,y)
107 if v=="?" then return v else i.lo = math.min(i.lo, v)
108 i.hi = math.max(i.hi, v)
109 i.y:add(y) end end
110
111 function RANGE.selects(i,t, v)
112 v = t.cells[i.at]
113 return v=="?" or (i.lo==i.hi and i.lo==v) or (i.lo<v and v<i.hi) end
114
115 function RANGE.score(i,goal,B,R, how, b,r,z)
116 how={}
117 how.more= function(b,r) return ((b<r or b+r < .05) and 0) or b^2/(b+r) end
118 how.less= function(b,r) return ((r<b or b+r < .05) and 0) or r^2/(b+r) end
119 how.tabu= function(b,r) return 1/(b+r) end
120 b, r, z = 0, 0, 1/big
121 for v,n in pairs(i.y.all) do
122 if v==goal then b = b+n else r=r+n end end
123 return how[the.How or "good"](b/(B+z), r/(R+z)) end
124
125 function NUM.new(i,at,txt)
126 i.at=at or 0; i.txt=txt or ""; i.w = wght(i.txt)
127 i.all,i.n,i.ok,i.lo,i.hi={}
128
129 function NUM.add(i,v)
130 if v == "?" then
131 i.lo=math.min(v,i.lo); i.hi=math.max(v,i.hi); push(i.all,v); i.ok=false end end
132
133 function NUM.norm(i,v)
134 return v=="?" and v or (i.hi-i.lo) < 1E-9 and 0 or (v-i.lo)/(i.hi-i.lo) end
135
136 function NUM.dist(i,v1,v2)
137 if v1=="?" and v2=="?" then return 0 end
138 if v1=="?" then v2=i:norm(v2); v1= v2<.5 and 1 or 0
139 elseif v2=="?" then v1=i:norm(v1); v2= v1<.5 and 1 or 0
140 else v1, v2 = i:norm(v1), i:norm(v2) end
141 return math.abs(v1-v2) end
142
143 function NUM.has(i) if not i.ok then sort(i.all) end; i.ok=true; return i.all end
144 function NUM.mid(i) return per(i:has(),.5) end
145 function NUM.div(i, a) a=i:has(); return (per(a,.9) - per(a,.1))/2.56 end
146 function NUM.bin(i,x, b) b=(i.hi-i.lo)/the.bins;return math.floor(x/b+0.5)*b end
147 function NUM.binsMerge(i,ranges,min, a,b,c,j,n,tmp,expand)
148 function expand(t)
149 if #t<2 then return {} end
150 for j=2,#t do t[j].lo=t[j-1].hi end
151 t[1].x.lo, t[#t].x.hi = -big,big
152 return t
153 end
154 j,n,tmp = 1,#ranges,{}
155 while j<#n do
156 a, b = ranges[j], ranges[j+1]

```

```

156 if b then
157 c = a.y:merge(b.y,min)
158 if c then a = {lo=a.lo, hi=b.hi, y=c}
159 j = j+1 end end
160 tmp[#tmp+1] = a
161 j = j+1 end
162 return #tmp==#ranges and expand(tmp) or i:binsMerge(tmp,min) end
163
164 -----
165 function ROWS.new(i,src)
166 i.all, i.cols, i.xs, i.ys, i.names = {},(),(),(),nil
167 if type(src)=="table" then for _,r in pairs(src) do i:add(r) end
168 else for r in csv( src) do i:add(r) end end end
169
170 function ROWS.clone(i,init, j)
171 j=ROWS(i.names); for _,r in pairs(inits or {}) do j:add(r) end; return j end
172
173 function ROWS.add(i,t, r)
174 if i.names
175 then r = t.cells and t or ROW(i,t); i:update(r.cells); push(i.all, r)
176 else i:header(t) end end
177
178 function ROWS.header(i,t, col)
179 i.names = t
180 for at,txt in pairs(t) do
181 col = push(i.cols, (nump(txt) and NUM or SYM)(at,txt))
182 if not skip(txt) then push(goalp(txt) and i.ys or i.xs, col) end end end
183
184 function ROWS.update(i,t)
185 for _,col in pairs(i.cols) do col:add(t[col.at]) end end
186
187 function ROWS.around(i,r1,t, fun)
188 function fun(r2) return (dist=r1:dist(r2), row=r2) end
189 return sort(map(t or i.all, fun), lt"dist")
190
191 function ROWS.far(i,r1,t, tmp)
192 tmp = i:around(r1,t)
193 return tmp[(#tmp)*the.Far//1].row end
194
195 function ROWS.mid(i,cols) return map(cols or i.ys, function(col) return col:mid() end) end
196
197 function ROWS.lo(i,cols) return map(cols or i.ys, function(col) return col.lo end) end
198
199 function ROWS.look(i, w,sample,best,rests)
200 w = i.all
201 sample = many(w, the.Some)
202 rests = {}
203 best = i:far(any(sample), sample)
204 for _,stop in pairs(({#w}*the.min,the.Min)) do
205 while #w > stop do
206 local rest = i:far(best, sample)
207 if rest < best then best,rest = rest,best end
208 best.evaluated, rest.evaluated = true,true
209 local c = best:dist(rest)
210 for _,r in pairs(w) do r.x=(r:dist(best)^2 + c^2- r:dist(rest)^2)/(2*c) end
211 local bests = {}
212 for n,r in pairs(sort(w,lt"x")) do push(n<=#w/2 and bests or rests,r) end
213 w=bests
214 sample = many(w,the.Some) end end
215 return ra,w,many(rests, #w*the.also) end
216
217 function ROWS.how(i, bests, rests)
218 local bins={}
219 for _,col in pairs(XXX.xs) do
220 for _,bin in pairs(ranges(col, bests, rests)) do
221 push(bins, {score=bin:score(i,#bests,#rests), bin=bin}) end end
222 for _,bin in pairs(sort(bins,gt"score")) do print(bin) end end
223
224 -----
225 return (NUM=NUM,ROWS=ROWS, ROW=ROW, help=help, the=the)

```