



```
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15 -- In this code:
16 -- - Line strive to be 80 chars (or less)
17 -- - Two spaces before function arguments denote optionals.
18 -- - Four spaces before function arguments denote local variables.
19 -- - Private functions start with '-'
20 -- - Arguments of private functions do anything at all
21 -- - Local variables inside functions do anything at all
22 -- - Arguments of public functions use type hints
23 -- - Variable 'x' is is anything
24 -- - Prefix 'is' is a boolean
25 -- - Prefix 'fun' is a function
26 -- - Prefix 'f' is a filename
27 -- - Prefix 'n' is a string
28 -- - Prefix 's' is a string
29 -- - Prefix 'c' is a column index
30 -- - 'col' denotes 'num' or 'sym'
31 -- - 'x' is anything (table or number of boolean or string
32 -- - 'v' is a simple value (number or boolean or string)
33 -- - Suffix 's' is a list of things
34 -- - Tables are 't' or, using the above, a table of numbers would be 'ns'
35 -- - Type names are lower case versions of constructors; e.g 'col' isa 'Cols'.
36 local l=require"lib"
37 local _=require"sam"
38
39 local o,oo,per,push,rnd = l.o,l.oo,l.per,l.push,l.rnd
40 local add,adds,dist,div = _._add,_._adds,_._dist,_._div
41 local mid, records, the = _._mid,_._records,_._the
42 local Num,Sym = _._Num, _._Sym
43
44 local eg= {}
45 function eg.the() oo(the); return true end
46
47 function eg.ent( sym,ent)
48 sym= adds(Sym()), {"a","a","a","a","b","b","c"})
49 ent= div(sym)
50 print(ent,mid(sym))
51 return 1.37 <= ent and ent <=1.38 end
52
53 function eg.num( num)
54 num=Num()
55 for i=1,100 do add(num,i) end
56 local med,ent = mid(num), rnd(div(num),2)
57 print(mid(num) ,rnd(div(num),2))
58 return 50<= med and med<= 52 and 30.5 <ent and ent <32 end
59
60 function eg.bignum( num)
61 num=Num()
62 the.nums = 32
63 for i=1,1000 do add(num,i) end
64 oo(_._nums(num))
65 return 32==#num._has end
66
67 function eg.read()
68 oo(records("./data/aut93.csv").cols.y); return true end
69
70 function eg.dist( data,t)
71 data=records("./data/aut93.csv")
72 t={}
73 for i=1,256 do push(t,rnd(dist(data,1.any(data.rows), 1.any(data.rows)),2)) end
74 table.sort(t)
75 oo(t)
76 return true end
77
78 -----
79 the = l.cli(the)
80 os.exit( l.run(the.eg, eg, the))
```

```
81
82
83
84
85
86 -- For a list of coding conventions in this file, see
87 -- [eg.lua] (https://github.com/timm/lua/blob/main/src/sam/eg.lua).
88 local l=require"lib"
89 local the,l.settings[{}
90 SAM : Semi-supervised And Multi-objective explanations
91 (c) 2022 Tim Menzies <tim@ieee.org> BSD-2 license
92
93 USAGE: lua eg.lua [OPTIONS]
94
95 OPTIONS:
96 -e --eg start-up example = nothing
97 -h --help show help = false
98 -n --nums how many numbers to keep = 256
99 -p --p distance coefficient = 2
100 -s --seed random number seed = 10019]]
101 -- Commonly used lib functions.
102 local o,oo,per,push = l.o,l.oo,l.per, l.push
103
104 ----- Classes
105 local Data,Cols,Sym,Num,Row
106 -- Holder of 'rows' and their summaries (in 'cols').
107 function Data() return {cols=nil, rows={}} end
108
109 -- Holder of summaries
110 function Cols(t) return {klass=nil,names={},nums={}, x={}, y={}, all={}} end
111
112 -- Summary of a stream of symbols.
113 function Sym(c,s)
114 return {n=0,at=c or 0, names=or "", _has={}} end
115
116 -- Summary of a stream of numbers.
117 function Num(c,s)
118 return {n=0,at=c or 0, names=or "", _has={},
119 isNum=true, lo= math.huge, hi= -math.huge, sorted=true,
120 w={s or ""}:find"-%S" and -1 or 1) end
121
122 -- Hold one record, in 'cells' (and 'cooked' is for discretized data).
123 function Row(t) return {cells=t, cooked=l.copy(t)} end
124
125 ----- Data Functions
126 local add,adds,clone,div,mid,norm,nums,record,records,stats
127 ----- Create
128 -- Generate rows from some 'src'. If 'src' is a string, read rows from file;
129 -- else read rows from a 'src' table. When reading, use rowl to define columns.
130 function records(src, data, oneRow,head)
131 data = data or Data()
132 function oneRow(t)
133 if data.cols then record(data,t) else data.cols=head(t) end end
134 function head(sNames)
135 local cols = Cols()
136 cols.names = sNames
137 for c,s in pairs(sNames) do
138 local col = push(cols.all, -- Numerics start with Uppercase.
139 (s:find"^[A-Z]" and Num or Sym)(c,s))
140 if not s:find"%" then -- some columns are skipped
141 push(s:find"[+-]" and cols.y or cols.x, col) -- some cols are goal cols
142 if s:find"$" then cols.klass=col end end end
143 return cols
144 end
145 if type(src)=="string" then l.csv(src, oneRow)
146 else for _,t in pairs(src or {}) do oneRow(t) end end
147 return data end
148
149 -- Return a new data with same structure as 'data1'. Optionally, oad in 'rows'.
150 function clone(data1, rows)
151 data2=Data()
152 data2.cols = _._head(data1.cols.names)
153 for _,row in pairs(rows or {}) do record(data2,row) end
154 return data2 end
155
156 ----- Update
157 -- Add one thing to 'col'. For Num, keep at most 'nums' items.
158 function add(col,v)
159 if v=="?" then
160 col.n = col.n + 1
161 if not col.isNum then col._has[v] = 1 + (col._has[v] or 0) else
162 col.lo = math.min(v, col.lo)
163 col.hi = math.max(v, col.hi)
164 local pos
165 if #col._has < the.nums then pos = 1 + (#col._has)
166 elseif math.random() < the.nums/col.n then pos = math.random(#col._has) end
167 if pos then col.sorted = false
168 col._has[pos] = tonumber(v) end end end
169
170 -- Add many things to col
171 function adds(col,t) for _,v in pairs(t) do add(col,v) end; return col end
172
173 -- Add a new 'row' to 'data'. Calls 'add()' to update the 'cols' with new values
174
175 function record(data,xs)
176 local row= push(data.rows, xs.cells and xs or Row(xs)) -- ensure xs is a Row
177 for _,todo in pairs(data.cols.x, data.cols.y) do
178 for _,col in pairs(todo) do
179 add(col, row.cells[col.at]) end end end
180
181 ----- Query
182 -- Return kept numbers, sorted.
183 function nums(num)
184 if not num.sorted then table.sort(num._has); num.sorted=true end
185 return num._has end
186
187 -- Normalized numbers 0..1. Everything else normalizes to itself.
188 function norm(col,n)
189 return x=="?" or not col.isNum and x or (n-col.lo)/(col.hi-col.lo + 1E-32) end
190
191 -- Diversity (standard deviation for Nums, entropy for Syms)
192 function div(col)
193 if col.isNum then local a=nums(col); return (per(a,.9)-per(a,.1))/2.58 else
194 local function fun(p) return p*math.log(p,2) end
195 for _,n in pairs(col._has) do if n>0 then e=e+fun(n/col.n) end end
196 return e end end
197
198 -- Central tendency (median for Nums, mode for Syms)
```

```
199 function mid(col)
200 if col.isNum then return per(nums(col),.5) else
201 local most,mode = -1
202 for k,v in pairs(col._has) do if v>most then mode=most=k,v end end
203 return mode end end
204
205 -- For 'showCols' (default='data.cols.x') in 'data', report 'fun' (default='mid').
206 function stats(data, showCols,fun, t)
207 showCols, fun = showCols or data.cols.y, fun or mid
208 t={}; for _,col in pairs(showCols) do t[col.name]=fun(col) end; return t end
209
210 ----- Distance functions
211 local dist
212 -- Distance between two rows (returns 0..1). For unknown values, assume max distan
213 ce.
214 function dist(data,t1,t2)
215 local function fun(col, v1,v2)
216 if v1=="?" and v2=="?" then return 1 end
217 if not col.isNum then return v1==v2 and 0 or 1 end
218 v1,v2 = norm(col,v1), norm(col,v2)
219 if v1=="?" then v1 = v2<.5 and 1 or 0 end
220 if v2=="?" then v2 = v1<.5 and 1 or 0 end
221 return math.abs(v1-v2)
222 end
223 local d = 0
224 for _,col in pairs(data.cols.x) do
225 d = d + fun(col, t1.cells[col.at], t2.cells[col.at])^the.p end
226 return (d/#data.cols.x)^(1/the.p) end
227
228 ----- That's all folks.
229 return {the=the,
230 Data=Data, Cols=Cols, Sym=Sym, Num=Num, Row=Row,
231 add=add, adds=adds, clone=clone, dist=dist, div=div,
232 mid=mid, nums=nums, records=records, record=record, stats=stats}
```

