

```
1 local help= []
2 NB:
3 (c)2022 Tim Menzies, timm@ieee.org
4
5 OPTIONS:
6 -k -k handle rare classes = 1
7 -m -m handle rare attributes = 2
8 -p -p distance coefficient = 2
9 -w -w wait before classifying = 5
10
11 OPTIONS (other):
12 -h -h help show help = false
13 -g -g start-up goal = nothing
14 -s -s seed seed = 10019
15 -f -f file file = ../data/auto93.csv]
```

```

18 local lib = require"lib"
19 local argmax = lib.argmax
20 local cli, csv, demos, is, normpdf = lib.cli, lib.csv, lib.demos, lib.is, lib.normpdf
21 local oo, push, read, rnd, str = lib.oo, lib.push, lib.read, lib.rnd, lib.str
22
23 THE={}
24 help:gsusb("[-|-|[-|([%s|)|^|n|)%s|(%s|)%",function(key,x) THE[key] = read(x) end)
25
26 local NB,NUM,SYM,COLS,ROW,ROWS= is"NB",is"NUM",is"SYM",is"COLS",is"ROW",is"ROWS"
27
28 -- C O L S
29
30 function NUM.new(i) i.n,i.mu,i.m2,i.mu = 0,0,0,0 end
31 function NUM.mid(i,p) return rnd(i,mu,p) end
32 function NUM.like(i,x,...) return normpdf(x,i.mu,i.sd) end
33 function NUM.add(i,v, d)
34 if v=="?" then return v end
35 i.n = i.n + 1
36 d = v - i.mu
37 i.mu = i.mu + d/i.n
38 i.m2 = i.m2 + d*(v - i.mu)
39 i.sd = i.n<2 and 0 or (i.m2/(i.n-1))^0.5 end
40
41 function SYM.new(i) i.n,i.syms,i.most,i.mode = 0,{},0,nil end
42 function SYM.mid(i,...) return i.mode end
43 function SYM.like(i,x,prior) return ((i.syms[x] or 0)+THE.m*prior)/(i.n+THE.m) end
44 function SYM.add(i,v)
45 if v=="?" then return v end
46 i.n = i.n + 1
47 i.syms[v] = (inc or 1) + (i.syms[v] or 0)
48 if i.syms[v] > i.most then i.most,i.mode = i.syms[v],v end end
49
50 -- C O I S
51
52 local function usep(x) return not x:find"$" end
53 local function nump(x) return x:find"%A-Z" end
54 local function goalp(x) return x:find"%A-Z" end
55 local function klassp(x) return x:find"%I-S" end
56 local function new(at,txt)
57 txt = txt or ""
58 local i = (nump(txt) and NUM or SYM)()
59 i.txt,i.usep,i.at,i.w = txt,usep(txt),at or 0,txt:find"$" and -1 or 1
60 return i end
61
62 function COLS.new(i,t, col)
63 i.all,i.xs,i.ys,i.names = {},{}},{},t
64 for at,x in pairs(t) do
65 col = push(i.all, new(at,x))
66 if col.usep then
67 if klassp(col.txt) then i.klass=col end
68 push(golp(col.txt) and i.ys or i.xs, col) end end end
69
70 function COLS.add(i,t)
71 for _,col in pairs(i.xs,i.ys) do
72 for _,col in pairs(cols) do col:add(t[col.at]) end end
73 return t end
74
75 -- F O W V
76
77 function ROW.new(i,of,cells) i.of,i.cells,i.eval=of,cells,false end
78 function ROW.klass(i) return i.cells[i.of.cels.klass.at] end
79
80 -- F O W V S
81
82 local function load(src, fun)
83 if type(src)=="string" then for _,t in pairs(src) do fun(t) end
84 else for t in csv(src) do fun(t) end end end
85
86 function ROWS.new(i,t,i.cols=COLS(t); i.rows={} end
87 function ROWS.add(i,t)
88 t=.cells and t or ROW(i)()
89 i.cols:add(t.cells)
90 return push(i.rows, t) end
91
92 function ROWS.mid(i, cols, p, t)
93 t={};for _,col in pairs(cols or i.cols.ys) do t[col.txt]=col:mid(p) end;return t end
94
95 function ROWS.clone(i, f)
96 j= ROWS(i.cols.names);for _,row in pairs(t or {}) do j:add(row) end;return j end
97
98 function ROWS.like(i,t, nklases, nrows, prior,like,inc,x)
99 prior = (#t.cels + THE.k / (nrows + THE.k * nklases)
100 like = math.log(prior)
101 for _,col in pairs(i.cols.ys) do
102 x = t.cells[col.at]
103 if x and x ~= "" then
104 inc = col:like(x,prior)
105 like = like + math.log(inc) end end
106 return like end
107
108 -- F B
109
110 -- (0) Use row to initial our 'overall' knowledge of all rows.
111 -- After that (1) add row to 'overall' and (2) ROWS about this row's klass.
112 -- (3) After 'wait' rows, 'classify' row BEFORE updating training knowledge
113 function NB.new(i,src, all,nil,k)
114 i.overall,i.set,i.list = nil,{},{}
115 local function guess(row)
116 return argmax(i.set, function(x) return x:like(row,#i.list,#i.overall.rows) end)end
117
118 load(src,function(row, k)
119 if not i.overall then i.overall = ROWS(row) else -- (0) eat row!
120 row = i.overall:add(row) -- (1) add to overall
121 if #i.overall.rows > THE.wait then
122 print(row:klass(), guess(row)) end -- (3) classify before updating
123 k = row:klass() -- what klass is this?
124 i.set[k] = i.set[k] or push(i.list, i.overall:clone()) -- klass is known
125 i.set[k].txt = row.txt -- each klass knows its name
126 i.set[k]:add(row) end end -- (2) add to this row's klass

```

```

120 local no,go = {},{}
131 function go.the(t) return type(TH.p)=="number" and THE.p==2 end
133
134 function go.argmax(t,tfun)
135     local fun=function(x) return -x end
136     t={50,40,0,40,50}
137     return 3 == argmax(t,tfun) end
138
139 function go.num(n) n=NUM(i); for i=1,100 do n:add(i) end; return n.mu==50.5 end
140
141 function go.sym(s)
142     s=SYM(i)
143     for i in pairs{"a","a","a","a","b","b","b","c"} do s:add(x) end
144     return s.mode=="a" end
145
146 function go.csv(n,s)
147     n,s=0,0; for row in csv(TH.file) do n=n+1; if n>1 then s=s+row[1] end end
148     return rnd(s/n,3) == 5.441 end
149
150 function go.rows(rows)
151     load(TH.file,function(t) if rows then rows:add(t) else rows=ROWS[t] end end)
152     return rows.count.y[s[1].sd,0]==847 end
153
154 function go.nb(i)
155     return 268 == #NB(C.J./data/diabetes.csv)..set{"positive"}.rows end
156
157 -- s t c i r t -----
158
159 if pcall(debug.getlocal, 4, 1)
160 then return (ROW=ROW, ROWS=ROWS, NUM=NUM, SYM=SYM, THE=THE, lib=lib)
161 else THE = c1t(TH.help)
162     demos(TH,go) end

```