

```

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 --wait wait before classifying = 10
10
11 OPTIONS (other):
12 -h --help show help = false
13 -g --go start-up goal = nothing
14 -s --seed seed = 10019
15 -f --file file = ../data/auto93.csv]]
16
17 local lib = require"lib"
18 local cli, csv, demos, is, normpdf = lib.cli, lib.csv, lib.demos, lib.is, lib.normpdf
19 local oo, push, read, rnd, str = lib.oo, lib.push, lib.read, lib.rnd, lib.str
20
21 local THE={}
22 help:gsub("[^-][^%s+][^%n]*%s{[^%s+}]", function(key,x) THE[key] = read(x) end)
23
24 local NB, NUM, SYM, COLS, ROW, ROWS= is"NB", is"NUM", is"SYM", is"COLS", is"ROW", is"ROWS"
25
26
27
28
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
51
52 local function usep(x) return not x:find"$" end
53 local function nump(x) return x:find"[A-Z]" end
54 local function goalp(x) return x:find"[+-S]" end
55 local function klassp(x) return x:find"$" end
56
57 local function new(at, txt)
58 txt = txt or ""
59 local i = (nump(txt) and NUM or SYM)()
60 i.txt, i.usep, i.at, i.w = txt, usep(txt), at or 0, txt:find"$" and -1 or 1
61 return i end
62
63 function COLS.new(i, t, col)
64 i.all, i.xs, i.ys, i.names = {}, {}, {}, t
65 for at, x in pairs(t) do
66 col = push(i.all, new(at, x))
67 if col.usep then
68 if klassp(col.txt) then i.klass=col end
69 push(golp(col.txt) and i.ys or i.xs, col) end end end
70
71 function COLS.add(i, t)
72 for _, cols in pairs(i.xs, i.ys) do
73 for _, col in pairs(cols) do col:add(t[col.at]) end end
74 return t end
75
76
77
78 function ROW.new(i, of, cells) i.of, i.cells, i.evald=of, cells, false end
79 function ROW.klass(i) return i.cells[i.of.cols.klass.at] end
80
81
82
83 local function load(src, fun)
84 if type(src)=="string" then for _, t in pairs(src) do fun(t) end
85 else for t in csv(src) do fun(t) end end end
86
87 function ROWS.new(i, t) i.cols=COLS(t); i.rows={} end
88 function ROWS.add(i, t)
89 t.t.cells and t or ROW(i, t)
90 i.cols:add(t.cells)
91 return push(i.rows, t) end
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, t, j)
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, has)
99 prior = (i.n + THE.k) / (nrows + THE.k * nklases)
100 like = math.log(prior)
101 for _, col in pairs(i.cols.xs) do
102 x = t[col.at]
103 if x and x ~= "?" then
104 like = like + math.log(col:like(x, prior)) end end
105 return like end
106
107 function NB.new(i, src, all, one, kl)
108 i.all, i.one = nil, {}
109 load(src, function(t) if i.all
110 then kl = i.all:add(t):klass()
111 i.one[kl] = i.one[kl] or i.all:clone()
112 i.one[kl]:add(t)
113 else i.all = ROWS(t) end end) end

```

```

114
115
116
117 local no, go = {}, {}
118 function go.the() return type(THE.p)=="number" and THE.p==2 end
119 function go.num(n) n=NUM(); for i=1,100 do add(n,i) end; return n.mu==50.5 end
120
121 function go.sym(s)
122 s=SYM(); add(s, {"a","a","a","a","b","b","c"}); return s.mode=="a" end
123
124 function go.csv( n, s)
125 n,s=0,0; for row in csv(THE.file) do n=n+1; if n>1 then s=s+row[1] end end
126 return rnd(s/n, 3) == 5.441 end
127
128 function go.rows( rows)
129 load(THE.file, function(t) if rows then rows:add(t) else rows=ROWS(t) end end)
130 return rnd(rows.cols.ys[1].sd, 0)==847 end
131
132 function go.nb()
133 return 268 == #NB("../data/diabetes.csv").one.positive.rows end

```

```

134
135
136
137 if pcall(debug.getlocal, 4, 1)
138 then return (ROW=ROW, ROWS=ROWS, NUM=NUM, SYM=SYM, THE=THE, lib=lib)
139 else THE = cli(THE, help)
140 demos(THE, go) end

```