

```

1 local coerce, csv, fmt, goalp, lessp, nump, oo, o, sort, the
2 the=(file=".data/aaauto.cs", p=2)
3
4 function ignorep(s) return s:find"$" end
5 function klassp(s) return s:find"$" end
6 function lessp(s) return s:find"<" end
7 function goalp(s) return s:find"[>+]" end
8 function nump(s) return s:find"[A-Z]" end
9
10 function sort(t,f) table.sort(t,f) return t end
11 function lt(x) return function(a,b) return a[x] < b[x] end end
12
13 function per(t,p) return t[ ((p or .5)*#t) // 1 ] end
14
15 function coerce(x)
16     x = x:match("%s*(-)%s*$")
17     if x=="true" then return true elseif x=="false" then return false end
18     return math.tointeger(x) or tonumber(x) or x end
19
20 function csv(src)
21     src = io.input(src)
22     return function(line, row)
23         line=io.read()
24         if not line then io.close(src) else
25             row={} for k in line:gmatch("[^,]+") do row[1+#row]=coerce(x) end
26             return row end end end
27
28 fmt=string.format
29 function oo(t) print(o(t)) end
30 function o(t, u, one, sorted)
31     sorted = #t>0 -- true when array's indexes are 1,2,..#t
32     one= function(k,v) return sorted and tostring(v) or fmt("%.5s %s",k,v) end
33     u={} for k,v in pairs(t) do u[1+#u] = one(k,v) end
34     return (t.is or "").."["..table.concat(sorted and u or sort(u),",").."]" end
35
36 function rogues( ok)
37     for _,k in pairs( { "G", "VERSION", "arg", "assert", "collectgarbage",
38         "coroutine", "debug", "dofile", "error", "getmetatable", "io", "ipairs",
39         "load", "loadfile", "math", "next", "os", "package", "pairs", "pcall",
40         "print", "rawequal", "rawget", "rawlen", "rawset", "require", "select",
41         "setmetatable", "string", "table", "tonumber", "tostring", "type", "utf8",
42         "warn", "xpcall" }) do ok[k]=true end
43     for k,v in pairs(_ENV) do if not ok[k] then print("?",k, type(v)) end end end
44
45 function obj(name, t, new, str)
46     function new(kl,...) local x=setmetatable({},kl); kl.new(x,...); return x end
47     t = {__tostring=o, is=name or ""}; t.__index=t
48     return setmetatable(t, {__call=new}) end
49
50 function cells(i, rows, t)
51     t={}; for _,r in pairs(rows) do x=r.cells[i]; if ~=?" then t[1+#t]=x end end
52     return t end
53
54 function mode(t, ent, most, mode)
55     ent, most, mode = 0,0,nil
56     for _,x in pairs(t) do
57         t[x] = 1+(t[x] or 0)
58         if t[x] > most then most, mode = t[x], x end end
59     for _,n in pairs(t) do if n>0 then ent = ent - n/#t*math.log(n/#t,2) end end
60     return mode, ent end
61
62 function median(t) t=sort(t); return per(t.5), (per(t,.9)-per(t,.1))/2.56 end
63
64 Num=obj"Num"
65 function Num:new(pos,s)
66     self.pos, self.txt, self.lo, self.hi = pos or 0,s or "",1E32, -1E32
67     self.w = lessp(self.txt) and -1 or 1 end
68
69 function Num:add(x)
70     if x=="?" then return x end
71     self.lo = math.min(x,self.lo)
72     self.hi = math.max(x,self.hi) end
73
74 function Num:norm(x, lo,hi)
75     lo,hi= self.lo, self.hi
76     return x=="?" and x or hi-lo < 1E-9 and 0 or (x - lo)/(hi - lo) end
77
78 function Num:dist(x,y)
79     if x=="?" and y=="?" then return 1 end
80     if x=="?" then y = self:norm(y); x = y<.5 and 1 or 0
81     elseif y=="?" then x = self:norm(x); y = x<.5 and 1 or 0
82     else x,y = self:norm(x), self:norm(y) end
83     return math.abs(x - y) end
84
85 function Num:mid(rows) return median(cells(self.pos, rows)) end
86
87 -----
88 Sym=obj"Sym"
89 function Sym:new(pos,s) self.pos, self.txt= pos or 0,s or "" end
90 function Sym:add(x) return x end
91 function Sym:dist(x,y) return x=="?" and y=="?" and 1 or x==y and 0 or 1 end
92 function Sym:mid(rows) return mode(cells(self.pos, rows)) end
93
94 -----
95 Cols=obj"Cols"
96 function Cols:new(names, it, num, sym, col)
97     self.names=names
98     self.x, self.y, self.all, self.nums={}, {}, {}, {}
99     for pos, name in pairs(name) do
100         col = push(self.all, (nump(name) and Num or Sym)(pos, name))
101         if not ignorep(name) then
102             if klassp(name) then self.klass = col
103             push(goalp(name) and self.y or self.x, col) end end end
104
105 function Cols:add(t)
106     for _, col in pairs(self.all) do col:add(t[col.pos]) end; return t end
107
108 -----
109 Row=obj"Row"
110 function Row:new(data,t)
111     self.data, self.cells, self.evaluated = data, t, false end
112
113 function Row:__sub(other, d, inc)
114     d = 0
115     for _, col in pairs(data.cols.x) do
116         inc = col:dist(self.cells[col.pos], other.cells[col.pos])^the.p
117         d = d+ inc*the.p end
118     return (d / #data.cols.n) ^ (1/the.p) end
119
120 function Row:__lt(other, s1,s2,,e,y,a,b)
121     y= self.data.cols.y
122     s1, s2, e = 0, 0, math.exp(1)
123     for _, col in pairs(y) do
124         a = col:norm(self.cells[col.pos])
125         b = col:norm(other.cells[col.pos])
126         s1= s1 - e*(col.w * (a - b) / #y)
127         s2= s2 - e*(col.w * (b - a) / #y) end
128     return s1/#y < s2/#y end
129
130 -----
131 Rows=obj"Rows"
132 function Rows:new() self.rows, self.cols = {}, nil end
133
134 function Rows:add(t)
135     if self.cols
136     then push(self.rows, Row(self, self.cols:add(t)))
137     else self.cols = Cols(t) end end
138
139 function Rows:load(file)
140     for n, row in csv(the.file) do self:add(row) end
141     return self end
142
143 function Rows:around(r1, rows, t)
144     t={}; for _, r2 in pairs(rows or self.rows) do push(t, {row=r2, d= r1 - r2}) end
145     return sort(t, lt"d") end
146
147 function Rows:far(r1, rows)
148     return per(self:around(r1, rows), the.far).row end
149
150
151 rogues()

```