

Friday Week 8.

Expert Systems & Knowledge Acquisition: Practical 1

Chris Thornton -- February 6, 1989

-- Introduction --

The practical involves (a) applying a learning algorithm to some data so as to extract useful categories and (b) devising a mechanism which can exploit those categories in the generation of a simple expert system.

The main subtasks are as follows.

1. Select a domain and obtain (or construct) a body of data (some databases are provided below).
2. Implement a learning mechanism which can derive useful classes from the data.
3. Write a program which can take the class descriptions generated by the learning mechanism and produce a simple expert system.
4. Demonstrate that each link in the chain works; i.e. show that by applying the learning mechanism to the raw data you get useful class descriptions, and that by applying the expert-system generator to the class descriptions you get a useful expert system.

Obviously, the "expert system" is going to be very basic indeed. However, it should display some of the features of something like Mycin (e.g. query-the-user).

-- What to hand in --

- the code for the expert system generator and the learning mechanism. (Code which is not properly commented will be deemed unreadable.)

- annotated trace of the system in operation. (Obviously, The way in which you generate this trace will differ depending on which language you use.)

- an overview of the structure of your program

- an account of the arguments in favour of your particular implementation

- full descriptions of all unresolved problems

You can use any AI language for this exercise; e.g. Pop11, Prolog, Clisp.

NB. This file is /usr/local/poplocal/local/teach/course/eska/practical1

-- Flag data --

See (Forsyth and Rada 1986, chapter 3) for details.

vars data = [	Flag	Stars	Bars	Strip	Hues	Xers	Icon	Hans	Word	Num	Type
[Alabama	0	0	0	0	3	1	N	0	0	0	C
[Arkansas	29	0	0	0	3	0	N	0	1	0	C
[Connecticut	0	0	0	0	5	0	Y	0	4	0	U
[Delaware	0	0	0	0	6	0	Y	2	4	2	U
[Florida	0	0	0	0	6	1	Y	1	15	0	C
[Georgia	13	1	0	0	3	1	Y	0	3	1	C
[Illinois	0	0	0	0	6	0	Y	0	6	2	U
[Iowa	0	2	0	0	5	0	Y	0	10	0	U
[Louisiana	0	0	0	0	4	0	Y	0	4	0	C
[Maryland	0	12	0	0	4	0	Y	0	0	0	U
[Massachusetts	0	0	0	0	4	0	Y	1	6	0	U
[Mississippi	13	0	3	3	3	1	N	0	0	0	C
[New Hants	9	0	0	0	5	0	Y	0	7	1	U
[New Jersey	0	0	0	0	5	0	Y	2	3	1	U
[New York	0	0	0	0	6	0	Y	2	1	0	U
[N. Carolina	1	1	2	4	0	0	N	0	3	4	C
[Ohio	17	0	5	3	0	0	N	0	0	0	U
[Rhode Is.	13	0	0	3	0	0	Y	0	1	0	U
[S. Carolina	0	0	0	2	0	0	Y	0	0	0	C
[Tennessee	3	2	0	3	0	0	N	0	0	0	C
[Texas	1	1	2	3	0	0	N	0	0	0	C
[Virginia	0	0	0	5	0	0	Y	2	4	0	C
[Wisconsin	0	0	0	5	0	0	Y	2	2	1	U
[Washington	3	0	5	2	0	0	N	2	0	0	U

-- Signal peptide data --

Taken from

Gascuel, O. & Danchin, A. (1988). Data analysis using a learning program, a case study: an application of PLAGE to a biological sequence analysis. In Y. Kodratoff (Ed.), PROCEEDINGS OF THE EIGHTH EUROPEAN CONFERENCE ON

ARTIFICIAL INTELLIGENCE. London: Pitman Publishing.

"Signal peptides are segments of proteins, of 15-40 amino acids, positioned at the start (called amino terminal end) of exported proteins. They play an important role in cell life since they allow export or secretion of proteins through their surrounding membranes.

have been considered similar, structurally as well as functionally. But recent experiments have lead researchers to question this hypothesis. Our work was to extract significant differences between two classes of signal peptides coming from an eukaryote (H. sapiens) and a prokaryote (E. coli). Here are two examples:

```

MKKTAIAIAVALAGFATVAQA (1)
MKVMRTTVATVVAATLSMSAFSVFA (2)

```

Two E. coli signal peptides.

Each class contains all known signal peptides: 22 H. sapiens and 18 E. coli (see data below)." (p. 391)

Each letter labels a particular amino acid. A sublist thus describes a single signal peptide.

```

vars data = [
/* H. sapiens signal peptides */
[MATGSRSTSLLLAFGLLCLPWLQEGSA]
[MDYYRKYAAIPLVTLVFLHVLHS]
[MPGSRSTSLLLAFALLCLPWLQEGSA]
[MPRLFLFHLLEFCLLLNQFSRAVAA]
[MALWMRLPLLLALLALWGPDPAAA]
[MAAARLCLSLLLSTCVALLQPLLGAQG]
[MPSSVSWGILLLAGLCLVPVSLA]
[MALTFALLVALLVLSCKSSGSGVG]
[MTNKKLLQIALLLCFSTTLAS]
[MKYTSYILAFQLCIVLGSIG]
[MEFGLSWLFLVAILKGVQC]
[MDMRVLAQLLGLLLCFPGARC]
[MILDKALMLALGALTVMSPCGG]
[MVCLKLPGGSSLAALTVTLMVLSSRLAFA]
[MKAAVLTAVFLTGSQA]
[MYSNVIGTVTSSKRKVYLLSLLIGFWDCTC]
[MFSMRIVCLVLSVVGTAWT]
[MKRMVSWSPHKLKTMKHLILLLCVPLVKS]
[MPRSCSRSGALLALLQASMEVRG]
[MKWVESIFLIFLLNFESR]
[MKWVWALLLLAAWAAA]
[MEPWPLLLFSLCSAGLVLG]

/* E. coli signal peptides */
[MKATKLVLGAVILGSTLLAG]
[MSIQHFRVALIPFFAFCPLPVFA]
[MMITLRKLPLAVAVAGVSAQAMA]
[MKKTAAIAIAVALAGFATVAQA]
[MKVKVLSSLVLPALLVAGAANA]
[MKKSTLALVVMGIVASASVQA]
[MKKLLPILIGLSLGSFSSLSQA]
[MIKSVIA GAVAMAVVSF GVNN]
[MKQSTIALALLPLLFTPTVTKA]
[MKVMRTTVATVVAATLSMSAFSVFA]
[MKIKTGARILALSALTMMFSAASA]
[MKANAKTIIAGMIALAISHTAMA]
[MKKLALSLSLVAFSSATAAFA]
[MKKSLSVLKASVAVATLVPMLSFA]
[MKKTVLALSLLIGLGATAASYA]
[MKRNILAVIVPALLVAGTANA]
[MKKLLFAIPLVVPFYSHS]
[MNIKGKALLAGCIALAFSNMALA]
];

```

-- Medical data --

Each sublist corresponds to one patient (record). The 1st item (eg

The feature-names are as follows.

[sex age onset site present site aggravating factors relieving factors progress duration type severity nausea vomiting anorexia prev indigestion jaundice bowels micturition prev similar pain prev abdo surgery drugs for abdo pain mood colour abdo movement scar distension tenderness rebound guarding rigidity mass murphys bowel sounds rectal-vaginal tenderness]

The range for each feature is as follows.

```

[[1 2] [3 10] [11 23] [24 36] [37 42] [43 48] [49 51]
[52 55] [59 61] [63 64] [65 66] [67 68] [69 70] [71 72] [73 74]
[75 79] [80 84] [85 86] [87 88] [89 90] [91 93] [94 98] [99 101]
[102 103] [104 105] [106 118] [119 120] [121 122] [123 124]
[125 126] [127 128] [129 131] [132 136]]

```

The actual diseases are:

```

app appendicitis
gb gall bladder (cholecystitis)
ino intestinal obstruction
rc renal colic.

```

[root flag. components bot]

-1-

[stars bars strip hues xers icon hans word num type]

[Alabama arkansas connecticut ... Washington]

```

vars data = [
[app 2 6 19 26 37 43 51 53 60 63 65 67 69 72 74 75 80 86 87 90 91 94 99 103 105 108
120 122 124 126 128 129 134]
[app 1 5 19 26 37 43 51 55 60 64 65 67 69 72 74 75 80 86 88 90 92 96 99 103 105 115
119 121 123 126 128 130 133]
[app 1 4 15 26 37 43 50 54 60 63 65 67 69 72 74 77 81 86 88 90 91 96 99 103 105 108
119 121 124 126 128 129 133]
[app 1 5 19 26 37 43 51 55 60 64 66 68 69 71 74 75 80 86 88 90 91 96 99 103 105 108
119 121 124 126 128 129 133]
[app 1 5 14 26 37 43 51 55 60 64 66 68 70 72 74 75 80 85 88 90 91 94 99 103 105 108
120 121 124 126 128 129 134]
[app 1 4 15 25 37 0 50 53 61 64 65 67 69 72 74 75 80 86 88 90 92 94 99 103 105 110
120 122 124 126 128 129 0]
[app 2 9 13 26 38 48 51 54 60 63 65 67 69 72 74 76 80 86 88 90 92 94 99 103 105 108
120 121 124 126 128 129 136]
[app 1 6 15 29 37 43 50 53 60 64 66 67 70 72 74 77 80 85 88 90 91 96 99 103 105 111
120 121 124 126 128 129 134]
[app 2 5 15 30 37 48 50 52 60 63 65 67 69 72 74 75 81 86 88 90 91 95 99 103 105 112
119 122 124 125 128 129 136]
[app 2 4 19 26 37 47 51 54 60 63 65 68 70 72 74 75 80 86 88 90 92 94 99 103 105 108
119 121 124 126 128 129 136]
[app 1 5 16 26 37 47 49 53 60 63 65 68 70 72 74 75 80 85 87 90 91 96 99 103 105 108
119 122 124 126 128 129 133]
[app 1 4 19 26 37 43 51 54 0 63 66 67 69 0 0 76 80 86 88 90 91 96 99 103 105 108 120
122 124 126 128 129 133]
[app 2 4 13 26 37 43 50 52 60 63 65 67 70 71 74 76 80 85 88 90 91 94 99 103 105 108
120 121 124 126 128 129 133]
[app 1 4 16 29 37 48 51 53 60 64 65 68 69 72 74 76 80 85 88 90 91 95 99 103 105 108
119 121 124 126 128 129 136]
[gb 2 9 11 24 42 44 51 52 60 63 66 67 69 72 74 75 80 86 88 90 91 94 99 103 105 106
120 122 124 126 128 129 0]
[gb 1 8 15 24 39 48 49 53 59 63 65 67 70 72 74 75 80 85 87 89 91 94 99 102 105 110
122 124 126 128 129 136]
[gb 1 10 11 24 37 43 50 55 60 63 65 68 69 72 74 76 80 85 88 90 91 94 99 103 105 106
120 121 124 125 127 129 136]
[gb 2 10 15 28 42 43 51 52 60 64 65 67 69 72 74 75 80 86 88 89 92 94 99 103 105 0 120
122 124 126 128 129 136]
[gb 2 7 15 28 42 48 50 55 59 63 65 68 69 72 74 76 80 86 88 90 91 94 99 103 105 110
120 121 124 126 128 129 136]
[gb 1 7 11 24 39 48 49 53 60 63 65 67 70 71 74 75 80 85 88 90 91 94 99 103 105 106
120 122 124 126 127 129 136]
[gb 1 10 15 28 42 48 51 52 60 64 65 67 0 72 74 75 81 86 87 89 92 94 99 102 105 110
120 122 124 126 127 129 136]
[gb 1 7 21 34 41 47 50 53 60 64 66 67 69 72 74 75 80 86 87 90 92 94 99 102 105 112
119 121 124 126 127 129 136]
[gb 2 10 13 26 38 43 49 52 60 63 66 68 69 71 74 79 83 86 88 90 91 95 99 103 104 112
119 121 124 126 127 131 136]
[gb 1 10 16 28 42 48 51 52 60 64 65 67 69 71 74 75 80 86 88 89 91 95 99 103 105 110
120 121 124 126 127 129 136]
[gb 1 9 20 24 37 43 51 55 60 64 65 67 69 71 74 76 84 86 87 90 92 97 99 102 105 106
120 121 124 126 128 129 136]
[gb 2 9 11 25 42 43 51 55 60 63 65 67 69 71 74 76 80 86 87 89 92 94 99 102 105 106
120 122 124 126 127 129 136]
[gb 1 10 15 28 38 48 49 53 60 63 65 67 69 72 74 75 81 86 88 90 91 97 99 103 105 110
120 122 124 126 127 129 136]
[gb 2 10 20 0 42 48 49 55 61 63 65 67 69 72 74 77 80 86 88 90 91 94 99 103 105 118
120 122 124 126 128 129 136]
[gb 1 9 20 33 42 43 49 55 59 63 66 68 69 72 74 77 80 86 88 89 91 96 99 103 105 106
120 122 124 126 128 129 136]
[gb 2 8 15 28 42 48 50 52 60 64 65 67 70 71 74 78 80 85 88 89 91 94 99 103 105 110
122 124 126 128 129 136]
[ino 2 10 14 27 42 44 49 52 61 64 66 67 70 72 74 75 82 85 87 90 91 94 99 102 105 109
120 122 124 0 128 131 136]
[ino 1 10 16 29 37 48 51 55 61 64 65 67 70 72 74 76 80 86 88 90 93 94 100 103 105 111
120 0 124 125 128 131 135]
[ino 2 10 12 25 37 48 51 52 60 0 65 67 0 72 74 75 80 86 87 90 92 0 0 102 105 110 0 0
124 125 128 129 0]
[ino 2 10 19 33 42 48 50 55 60 63 65 67 69 72 74 76 80 86 88 90 91 95 99 103 104 0
120 122 124 126 128 129 136]
[ino 2 10 19 32 0 0 0 0 0 0 66 67 69 72 74 76 83 86 88 90 92 94 100 103 104 108 120
122 124 125 128 0]
[ino 1 10 19 32 37 43 51 55 61 63 65 67 69 72 74 76 80 85 88 90 91 94 99 103 105 108
120 121 124 125 128 129 136]
[ino 1 8 16 29 42 48 51 55 59 63 65 67 69 72 74 76 80 86 88 90 93 94 99 103 105 109
120 122 124 126 128 131 136]
[ino 2 10 16 29 37 43 50 54 61 63 65 67 69 72 74 76 80 86 87 90 92 94 101 102 104 111
120 122 124 126 128 131 136]
[ino 2 7 16 29 38 48 51 55 61 63 65 67 69 72 74 76 80 86 88 90 91 94 99 102 104 111
120 122 124 126 128 129 136]
[ino 1 10 19 32 42 44 49 54 61 63 65 67 69 72 74 76 80 85 88 90 91 94 99 103 105 111
120 122 124 125 128 131 136]
[ino 2 10 19 32 42 48 50 55 60 63 65 67 69 71 73 76 80 86 87 90 91 94 101 102 104 110
120 122 124 125 128 131 136]
[ino 2 10 19 30 40 47 51 55 61 63 65 67 69 72 74 76 80 86 88 89 91 94 99 103 104 108
120 122 124 126 128 131 136]
[ino 1 6 20 33 37 47 49 54 59 63 65 67 69 71 74 76 82 85 88 90 92 94 100 103 104 0
120 122 124 126 128 131 0]

```

```

[ino 2 10 19 32 39 48 50 55 60 63 65 67 69 71 74 76 80 85 88 90 91 94 99 103 104 118
120 122 124 126 128 131 136]
[ino 1 10 19 32 42 48 50 53 60 63 66 68 70 72 74 75 80 85 88 90 91 94 99 103 104 0
120 122 124 126 128 129 135]
[ino 2 10 13 26 42 44 51 54 61 63 65 67 69 71 74 76 80 86 87 90 91 95 99 102 104 115
120 122 124 125 128 131 136]
[ino 1 6 15 31 37 43 51 53 61 64 65 67 69 72 74 75 80 86 87 90 92 95 99 102 104 113
120 122 124 126 128 131 136]
[ino 2 10 16 29 42 48 50 54 61 63 65 67 69 72 74 76 80 86 87 90 91 94 99 102 104 111
120 122 124 126 128 129 0]
[rc 2 6 17 30 42 43 51 52 60 64 66 67 69 72 74 75 80 85 88 89 92 94 99 103 105 112
119 121 124 126 128 129 0]
[rc 2 8 22 35 42 48 50 52 61 63 65 67 70 72 74 75 80 86 88 89 92 94 99 103 105 118 0
122 124 126 128 0]
[rc 1 9 22 35 38 48 50 53 60 64 65 67 69 72 74 75 80 86 88 89 92 94 99 103 105 117
120 121 124 126 128 129 136]
[rc 1 7 18 31 42 48 51 55 61 64 66 68 70 72 74 75 81 85 88 90 93 94 99 103 105 113
120 122 124 126 128 129 136]
[rc 1 8 13 26 42 47 51 53 61 64 65 67 69 71 73 75 80 86 87 89 92 94 99 102 105 118
120 122 124 126 128 129 136]
[rc 1 10 13 30 42 48 50 52 60 63 65 67 69 71 74 75 81 86 87 89 92 94 100 0 0 112 119
121 123 126 128 131 136]
];

```

-- Info. for BIRDS/Poplog users -----

Note that sourcing `/usr/local/poplocal/poplogin` in your `.login` now sets everything up for Poplog use. Sourcing `/usr/local/poplocal/popuclogin` does the same thing but arranges for Vcd to behave like MicroEMACS by default.

In Poplog, this file is accessible as

TEACH COURSE/ESKA/PRACTICAL1