Another Machine Learning Fable

Norman Fenton, 23 March 2015

A bank has a massive database on people who have had loans. The database looks like this:

Customer	Age	Marital status	Employment status	Home owner	Salary	Loan	…	Defaulted
1	37	M	Employed	Y	50000	10000	…	N
2	45	M	Self-employed	Y	60000	5000	…	N
3	26	M	Self-employed	Y	30000	20000	…	Y
4	29	S	Employed	N	50000	15000	…	N
5	26	M	Employed	Y	90000	20000	…	N
6	35	S	Self-employed	N	70000	20000	…	Y
7	32	M	Self-employed	Y	40000	5000	…	N
8	37	M	Employed	Y	25000		…	Y
9	18	S	Unemployed	N	0	50000	…	N
10	40	M	Employed	Y	65000	45000	…	N
11	21	S	Employed	N	20000	10000	…	Y
12	30	S	Employed	N	40000	5000	…	N
13	22	M	Self-employed	N	30000	10000	…	Y
14	35	M	Unemployed	Y	0	3000	…	Y
15	19	S	Unemployed	N	0	100000	…	N
…	…	…	…	…	…	…	…	…
100001	34	M	Employed	Y	45000	1000	…	N
100002	28	S	Self-employed	N	25000	2000	…	N
100003	19	S	Unemployed	N	0	25000	…	N
…	…	…	…	…	…	…	…	…

Because too many people 'default' on loans the bank wants to use machine learning techniques on this database to help decide whether or not to offer credit to new applicants. In other words they expect to 'learn' when to refuse loans on the basis that the customer profile is too 'risky'.

The fundamental problem with such an approach is that it can learn nothing about those customers who were refused credit precisely because the bank decided they were likely to default. Any causal knowledge about such (potential) customers is missing from the data.

Suppose, for example, that the bank normally refuses credit to people under 20, unless their parents are existing high-income customers known to a bank manager. Such special cases (like customers 9, 15, 100003 above) show up in the database and they never default. Any pure data driven learning algorithm will 'learn' that unemployed people under 20 never default - the exact opposite of reality in almost all cases. Pure machine learning will therefore recommend giving credit to people known most likely to default.

See also: Fenton NE, "A simple story illustrating why pure machine learning (without expert input) may be doomed to fail and totally unnecessary", 12 Nov 2012, http://www.eecs.qmul.ac.uk/~norman/papers/ml_simple_example.pdf