

# Neural Networks

Could we actually file one?



Christopher Cooksey, FCAS, MAAA, CSPA  
Head Actuary, Data & Analytics

20 March 2018



---

---

---

---

---

---

---

---

## Agenda

Brainstorming – Of course we can't
A Regulator's Perspective
How Neural Networks Work
The "Black Box"
Real Issues, Problems and Solutions

Page 2

---

---

---

---

---

---

---

---

## Brainstorming – Of course we can't

---

---

---

---

---

---

---

---

### Conventional Wisdom Says...

We could never use a Neural Network for a rating algorithm because...



---

---

---

---

---

---

---

---

### How Neural Networks Work

---

---

---

---

---

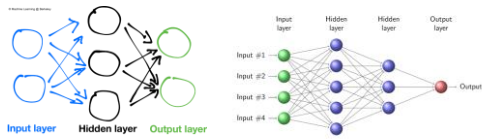
---

---

---

### Neural Networks

Artificial Neural Networks come in many varieties...



---

---

---

---

---

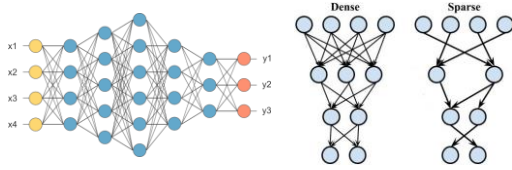
---

---

---

### Neural Networks

Artificial Neural Networks come in many varieties...



---

---

---

---

---

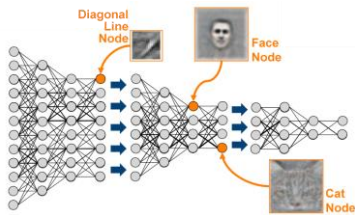
---

---

---

### Neural Networks

Artificial Neural Networks come in many varieties...



---

---

---

---

---

---

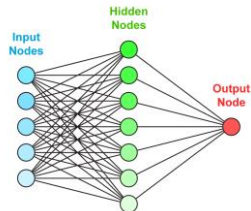
---

---

### Neural Networks

...so let's simplify our discussion.

A single layer, feedforward, backpropagation neural network with one output node.



---

---

---

---

---

---

---

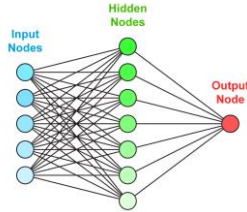
---

### Neural Networks

A single layer, feedforward, backpropagation neural network with one output node.

Input nodes are the predictor fields. Think design matrix:

- Each input node for a categorical predictor is one level, and the values are 0 or 1.
- Each input node for an ordinal predictor is the normalized value of that predictor.



Page 11

---

---

---

---

---

---

---

---

---

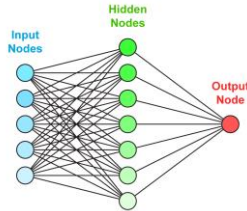
---

### Neural Networks

A single layer, feedforward, backpropagation neural network with one output node.

Each hidden node is a linear predictor, literally just a weighted combination of the input (i) for a given hidden node (j).

$$\mu_j = \sum_i (Input_i)w_{ij}$$



Page 12

---

---

---

---

---

---

---

---

---

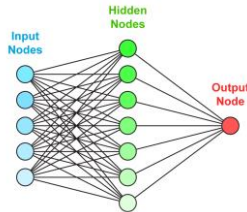
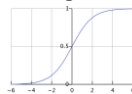
---

### Neural Networks

A single layer, feedforward, backpropagation neural network with one output node.

Except...the output from the linear predictor,  $\mu_j$ , is put through an "activation" function, such as a logistic function...

$$f(x) = \frac{1}{1 + e^{-x}}$$



Page 13

---

---

---

---

---

---

---

---

---

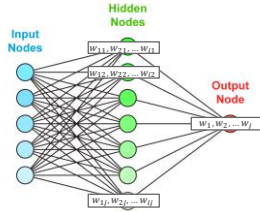
---

### Neural Networks

A single layer, feedforward, backpropagation neural network with one output node.

Finally, the transformed output from each hidden node is weighted together into the output.

Just like GLMs, we're trying to find optimal weights.



Page 11

---

---

---

---

---

---

---

---

---

---

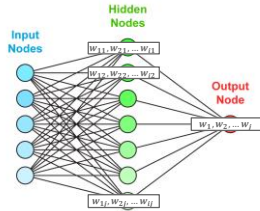
### Neural Networks

A single layer, feedforward, backpropagation neural network with one output node.

GLM optimal betas are found through minimizing a function on all the data.

Neural Net optimal weights are adjusted one record at a time until the error can't be reduced.

Each hidden node has different starting weights as well!



Page 11

---

---

---

---

---

---

---

---

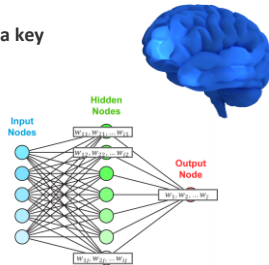
---

---

### Neural Networks

The activation function is also a key element of Neural Networks.

Just like the brain has different neurons fire for different tasks, each hidden node is sensitive to only certain inputs.



Page 11

---

---

---

---

---

---

---

---

---

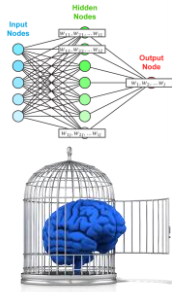
---

### Neural Networks

Finally, for purposes of this discussion, we will not focus on getting a good performing Neural Network.

We will assume we've created one for the purposes of rating and that it works well.

The question in this presentation is, how would we file it?




---

---

---

---

---

---

---

---

## The "Black Box"

### The "Black Box"

What do people mean when they say that Neural Networks are Black Boxes?

- You don't get to know what is happening in the hidden layer.
- Others?




---

---

---

---

---

---

---

---

### The "Black Box"

What makes neural networks harder to explain is that we can give no simple meaning to the weights.

However, the following statements are not necessarily true.

- Neural networks are complex.

Neural networks, just like GLMs or ensembles or nearly any other technique, will be as simple or complex as the modeler chooses to make them. This picture to the right would have 42 separate coefficients estimated.



Page 11

---

---

---

---

---

---

---

---

---

---

### The "Black Box"

What makes neural networks harder to explain is that we can give no simple meaning to the weights.

However, the following statements are not necessarily true.

- Neural networks are complex.
- Neural networks are not transparent in their operation.

Neural networks, just like GLMs, are a fairly straight-forward formula. How they operate is completely transparent. It is only the meaning of the individual components that is difficult.



Page 11

---

---

---

---

---

---

---

---

---

---

### The "Black Box"

What makes neural networks harder to explain is that we can give no simple meaning to the weights.

However, the following statements are not necessarily true.

- Neural networks are complex.
- Neural networks are not transparent in their operation.
- Nothing can be said about individual predictors in a neural network.

Not true! There are different ways to tackle this – relative importance of predictors, reason codes, observed versus modeled (OvM) charts, etc.



Page 11

---

---

---

---

---

---

---

---

---

---

# Real Issues, Problems and Solutions

---

---

---

---

---

---

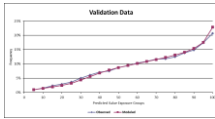
---

---

## Does the model perform well?

Does the model differentiate between customers (lift)?  
Does the model predict hold-out data?

It is easy enough to look at lift charts and OvM (observed versus modeled) charts to establish model performance.



Page 11

---

---

---

---

---

---

---

---

## Is the model stable over time?

Are the predictions relatively stable year over year?

This is an important question. Lift curves found individually for each year of experience can answer it.



Page 12

---

---

---

---

---

---

---

---



### Can we specify a rating plan? Calculate a rate?

Neural networks are a formula with weights. Not too hard to express; not too hard to calculate.

=1/(1+EXP(-SUMPRODUCT(B17:H17,B54:H54)-J17))

Example Policy													
	Age	# Drivers	Look Only	W	Look Only	W	Mult. Auto	W	Mult. Auto	W	Prof. Tesora	Year Age	
Values	42	2	0	0	0	0	1.00	1.00	1.00	0.00	0.00	1.00	
Standard Values	0.50	0.50	0.50	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	
Estimate:	0.4679												
Hidden Layer													
	Age	# Drivers	Look Only	W	Look Only	W	Mult. Auto	W	Mult. Auto	W	Prof. Tesora	Year Age	Weights
Node 1	-0.45	-0.80	0.50	-0.20	0.50	-0.50	-0.20	0.10	-0.50	0.50	0.00	-0.50	0.50
Node 2	0.50	-0.50	1.00	-1.00	-0.50	0.50	0.00	0.50	0.00	0.00	0.00	0.00	0.50
Node 3	0.70	0.50	-0.50	0.20	-0.20	0.20	0.20	-0.70	0.50	1.00	0.00	0.00	0.00
Node 4	0.50	-0.50	0.50	-0.50	-0.20	-0.20	0.20	0.00	-0.50	0.50	0.00	-0.50	0.00
Node 5	0.20	-0.50	0.00	-0.20	-0.20	0.20	0.20	0.00	0.50	1.00	0.00	-0.50	0.00
Node 6	-0.50	-0.50	0.50	-0.27	-0.24	-0.23	-0.24	-0.24	0.24	0.24	0.23	0.24	-0.23
Node 7	0.50	0.50	0.00	-0.50	-0.50	-0.50	-0.50	-0.50	0.50	0.50	0.50	0.50	0.50
Node 8	0.50	0.20	0.50	-0.50	-0.50	-0.22	-0.24	-0.22	-0.24	-0.22	-0.24	-0.22	-0.24
Node 9	-0.50	-0.50	-0.50	0.50	0.50	0.50	0.50	0.50	-0.50	-0.50	-0.50	-0.50	0.50
Node 10	-0.50	-0.50	-0.50	0.50	0.50	0.50	0.50	0.50	-0.50	-0.50	-0.50	-0.50	0.50

Page 11

### Can we specify a rating plan? Calculate a rate?

Neural networks are a formula with weights. Not too hard to express; not too hard to calculate.

=SUMPRODUCT(N16:N25,K16:K25)+M16

Example Policy													
	Age	# Drivers	Look Only	W	Look Only	W	Mult. Auto	W	Mult. Auto	W	Prof. Tesora	Year Age	
Values	42	2	0	0	0	0	1.00	1.00	1.00	0.00	0.00	1.00	
Standard Values	0.50	0.50	0.50	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	1.00	
Estimate:	0.5000												
Hidden Layer													
	Age	# Drivers	Look Only	W	Look Only	W	Mult. Auto	W	Mult. Auto	W	Prof. Tesora	Year Age	Weights
Node 1	-0.45	-0.80	0.50	-0.20	0.50	-0.50	-0.20	0.10	-0.50	0.50	0.00	-0.50	0.50
Node 2	0.50	-0.50	1.00	-1.00	-0.50	0.50	0.00	0.50	0.00	0.00	0.00	0.00	0.50
Node 3	0.70	0.50	-0.50	0.20	-0.20	0.20	0.20	-0.70	0.50	1.00	0.00	0.00	0.00
Node 4	0.50	-0.50	0.50	-0.50	-0.20	-0.20	0.20	0.00	-0.50	0.50	0.00	-0.50	0.00
Node 5	0.20	-0.50	0.00	-0.20	-0.20	0.20	0.20	0.00	0.50	1.00	0.00	-0.50	0.00
Node 6	-0.50	-0.50	0.50	-0.27	-0.24	-0.23	-0.24	-0.24	0.24	0.24	0.23	0.24	-0.23
Node 7	0.50	0.50	0.00	-0.50	-0.50	-0.50	-0.50	-0.50	0.50	0.50	0.50	0.50	0.50
Node 8	0.50	0.20	0.50	-0.50	-0.50	-0.22	-0.24	-0.22	-0.24	-0.22	-0.24	-0.22	-0.24
Node 9	-0.50	-0.50	-0.50	0.50	0.50	0.50	0.50	0.50	-0.50	-0.50	-0.50	-0.50	0.50
Node 10	-0.50	-0.50	-0.50	0.50	0.50	0.50	0.50	0.50	-0.50	-0.50	-0.50	-0.50	0.50

Page 11

### Again, the "rating plan"?

The rating plan is now a collection of input characteristics and specified weights. There is no Age curve in the same sense as the GLM.

Let's look at the example Rating Manual Handout.

#### Thoughts?

Defining minimums and maximums may be an issue.

Need to account for expenses too.

Page 11

### Which fields are more important?

Once the model is specified, it is a straight-forward task to calculate various metrics around variable importance.

- Replace all values of a field with its mean; refit the network. The loss of accuracy for each field quantifies its importance.
- For a single-layer network, you can take the absolute value of all the weights.

This question has been kicked around a lot. You can get an answer.

Page 21

---

---

---

---

---

---

---

---

---

---

### How do different fields behave in the model?

Observed versus Modeled (OvM) charts can give insight into the model. Just let the x-axis be the levels of the field instead of some sort by model prediction.

Don't forget that reality doesn't change just because we pick a different modeling technique.

If the severity increases as limit increases, then an OvM from any model should show this.



Page 22

---

---

---

---

---

---

---

---

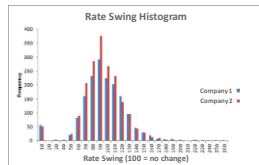
---

---

### Getting more practical – Rate Swings?

Rate swings are always an issue. Calculating them is easy. Controlling the size of them? Not so easy.

Is this different than a GLM though?



Page 23

---

---

---

---

---

---

---

---

---

---

### Getting more practical – Reversals?

How about reversals? No guarantees that there won't be reversals across certain fields.

Is this different than a GLM though?



Um, yeah. GLMs are easier for managing reversals...  
...except when using interactions and fields show up in multiple places. Then GLMs can be hard to manage reversals too.

Page 11

---

---

---

---

---

---

---

---

### More practicalities – FCRA?

Compliance with the Fair Credit Reporting Act is one of those things that we need to provide for even if not included in a filing.

Rates can be calculated with the old and new credit. If the rates increase, well, then you know.



Page 12

---

---

---

---

---

---

---

---

### More practicalities – Customer complaints?

Whenever rates change, there can be questions. Typically we need to provide reasons for a rate.

Age	MI	PO	MP
15	3.850	3.850	3.992
16	3.780	3.870	3.953
17	3.740	3.770	3.850
18	3.700	3.690	3.771
19	3.660	3.640	3.744
20	3.720	3.580	3.671
21	3.780	3.520	3.598
22	3.840	3.460	3.525
23	3.900	3.400	3.452
24	3.960	3.340	3.379
25	4.020	3.280	3.306
26	4.080	3.220	3.233
27	4.140	3.160	3.160
28	4.200	3.100	3.087
29	4.260	3.040	3.014
30	4.320	2.980	2.941
31	4.380	2.920	2.868
32	4.440	2.860	2.795
33	4.500	2.800	2.722
34	4.560	2.740	2.649
35	4.620	2.680	2.576
36	4.680	2.620	2.503
37	4.740	2.560	2.430
38	4.800	2.500	2.357
39	4.860	2.440	2.284
40	4.920	2.380	2.211
41	4.980	2.320	2.138
42	5.040	2.260	2.065
43	5.100	2.200	1.992
44	5.160	2.140	1.919
45	5.220	2.080	1.846
46	5.280	2.020	1.773
47	5.340	1.960	1.700
48	5.400	1.900	1.627
49	5.460	1.840	1.554
50	5.520	1.780	1.481
51	5.580	1.720	1.408
52	5.640	1.660	1.335
53	5.700	1.600	1.262
54	5.760	1.540	1.189
55	5.820	1.480	1.116
56	5.880	1.420	1.043
57	5.940	1.360	0.970
58	6.000	1.300	0.897
59	6.060	1.240	0.824
60	6.120	1.180	0.751
61	6.180	1.120	0.678
62	6.240	1.060	0.605
63	6.300	1.000	0.532
64	6.360	0.940	0.459
65	6.420	0.880	0.386
66	6.480	0.820	0.313
67	6.540	0.760	0.240
68	6.600	0.700	0.167
69	6.660	0.640	0.094
70	6.720	0.580	0.021
71	6.780	0.520	-0.052
72	6.840	0.460	-0.125
73	6.900	0.400	-0.198
74	6.960	0.340	-0.271
75	7.020	0.280	-0.344
76	7.080	0.220	-0.417
77	7.140	0.160	-0.490
78	7.200	0.100	-0.563
79	7.260	0.040	-0.636
80	7.320	-0.020	-0.709
81	7.380	-0.080	-0.782
82	7.440	-0.140	-0.855
83	7.500	-0.200	-0.928
84	7.560	-0.260	-1.001
85	7.620	-0.320	-1.074
86	7.680	-0.380	-1.147
87	7.740	-0.440	-1.220
88	7.800	-0.500	-1.293
89	7.860	-0.560	-1.366
90	7.920	-0.620	-1.439
91	7.980	-0.680	-1.512
92	8.040	-0.740	-1.585
93	8.100	-0.800	-1.658
94	8.160	-0.860	-1.731
95	8.220	-0.920	-1.804
96	8.280	-0.980	-1.877
97	8.340	-1.040	-1.950
98	8.400	-1.100	-2.023
99	8.460	-1.160	-2.096
100	8.520	-1.220	-2.169

Model Year	Rate
2010	3.850
2011	3.870
2012	3.890
2013	3.910
2014	3.930
2015	3.950
2016	3.970
2017	3.990
2018	4.010
2019	4.030
2020	4.050
2021	4.070
2022	4.090
2023	4.110
2024	4.130
2025	4.150
2026	4.170
2027	4.190
2028	4.210
2029	4.230
2030	4.250
2031	4.270
2032	4.290
2033	4.310
2034	4.330
2035	4.350
2036	4.370
2037	4.390
2038	4.410
2039	4.430
2040	4.450
2041	4.470
2042	4.490
2043	4.510
2044	4.530
2045	4.550
2046	4.570
2047	4.590
2048	4.610
2049	4.630
2050	4.650
2051	4.670
2052	4.690
2053	4.710
2054	4.730
2055	4.750
2056	4.770
2057	4.790
2058	4.810
2059	4.830
2060	4.850
2061	4.870
2062	4.890
2063	4.910
2064	4.930
2065	4.950
2066	4.970
2067	4.990
2068	5.010
2069	5.030
2070	5.050
2071	5.070
2072	5.090
2073	5.110
2074	5.130
2075	5.150
2076	5.170
2077	5.190
2078	5.210
2079	5.230
2080	5.250
2081	5.270
2082	5.290
2083	5.310
2084	5.330
2085	5.350
2086	5.370
2087	5.390
2088	5.410
2089	5.430
2090	5.450
2091	5.470
2092	5.490
2093	5.510
2094	5.530
2095	5.550
2096	5.570
2097	5.590
2098	5.610
2099	5.630
2100	5.650

For a GLM, just look at the relativities that the policy gets.

Age makes this policy's rate higher...  
Model Year makes this policy's rate lower...

Neural Network weights don't work this way!

Page 13

---

---

---

---

---

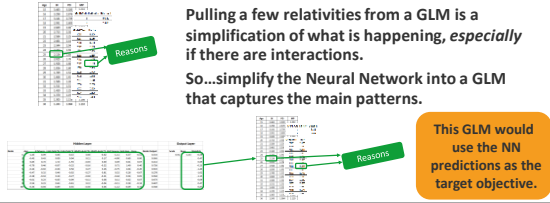
---

---

---

### More practicalities – Customer complaints?

Whenever rates change, there can be questions. Typically we need to provide reasons for a rate.




---

---

---

---

---

---

---

---

---

---

### Getting creative

It may be that this is all too much. The simplicity of the GLM approach in how it goes on to be implemented may be worth it.



You *will* lose something in the translation of the NN-based result to a GLM structure.

---

---

---

---

---

---

---

---

---

---

### Summary

- Unlike GLMs, the weights for neural networks do not have any straight-forward meaning.
- Nevertheless, like GLMs their operation is a well-defined formula.
- It is no harder to prove that a neural network is predicting well than any other model.
- It is still possible to get insights into how the model is working in relation to its predictors.
- Practical issues like reversals, rate swings and customer complaints are real issues that need to be, but can be, accounted for.

---

---

---

---

---

---

---

---

---

---

Questions?

Christopher Cooksey, FCAS, MAAA, CSPA  
Head Actuary, Data and Analytics

Guidewire Software  
ccooksey@guidewire.com

---

---

---

---

---

---

---

---