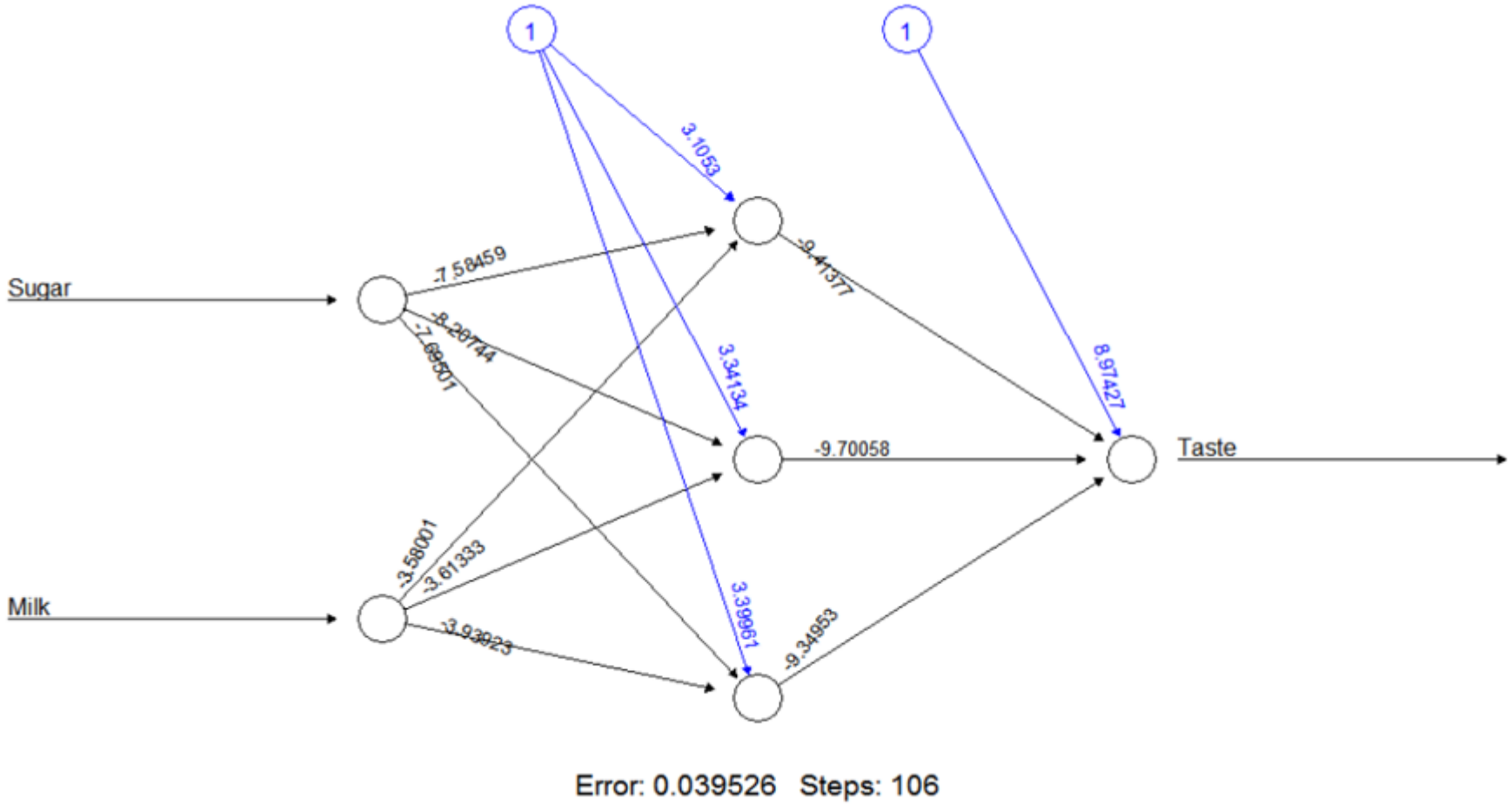
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Answers in Neural Network with R

BC2407 ANALYTICS II SEMINAR 5

NEUMANN CHEW C. H.

plot(ctt.m1) to *view the final weights on the Neural Network diagram*



Source: Textbook ADA Vol 1, Chap 9, Figure 9.2.

Q: What are the activation functions in each hidden node and output node?

1st hidden node:

$z_1 = 3.1053 - 7.58459(\text{sugar}) - 3.58001(\text{milk})$

Activation function: $f(z_1) = \frac{1}{(1+e^{-z_1})}$

2nd hidden node:

$z_2 = 3.34134 - 8.20744(\text{sugar}) - 3.61333(\text{milk})$

Activation function: $f(z_2) = \frac{1}{(1+e^{-z_2})}$

3rd hidden node:

$z_3 = 3.39961 - 7.69501(\text{sugar}) - 3.93923(\text{milk})$

Activation function: $f(z_3) = \frac{1}{(1+e^{-z_3})}$

Output node:

$z_4 = 8.97427 - 9.41377[f(z_1)] - 9.70058[f(z_2)] - 9.34953[f(z_3)]$

Activation function: $f(z_4) = \frac{1}{(1+e^{-z_4})}$

Here, the Y variable (i.e. Taste) is a binary variable. What would change if Y is (a) continuous or (b) multi-categorical instead of binary?

Ans:

The activation function $f(z_4)$ will be (a) Linear or (b) softmax (aka multinomial logistic function).

Refer to BC2406 logistic regression or Textbook Vol 1. Chap 7.

ctt.m1$net.result shows the model predicted value for each cases in the trainset

```
[[1]]
              [,1]
[1,] 9.863232e-01
[2,] 7.175456e-08
[3,] 1.006381e-03
[4,] 1.376937e-02
[5,] 9.900065e-01
[6,] 9.993707e-01
[7,] 9.997900e-01
```

- Q: What's the meaning of these numbers?
- Ans: Logistic function values that serve as P(Y = 1) for each of the 7 cases in the trainset.

Class Activity 1

Neural Network Interpretation

Est. Duration: 30 mins

1. The Chocolate Taste Test dataset, randomised start weights and final optimized weights in the neural network above are provided in Excel file CTT.xlsx.
2. For each set of weights, compute the activation functions and cross entropy error in Excel. Verify that the mean CE error is smaller using the final weights. What is the meaning of error in R output?

   Ans: The error in R output is the total CE error. See calculations in CTT solution.xlsx.
3. Verify the R output in ctt.m1$net.result against your excel calculations.

```
                          [,1]
error                0.039526475
reached.threshold    0.009733772
steps              106.000000000
Intercept.to.1layhid1  3.105303556
Sugar.to.1layhid1     -7.584593851
Milk.to.1layhid1      -3.580013512
Intercept.to.1layhid2  3.341337422
Sugar.to.1layhid2     -8.207442897
Milk.to.1layhid2      -3.613327958
Intercept.to.1layhid3  3.399614927
Sugar.to.1layhid3     -7.695007496
Milk.to.1layhid3      -3.939234882
Intercept.to.Taste     8.974265963
1layhid1.to.Taste     -9.413767383
1layhid2.to.Taste     -9.700584522
1layhid3.to.Taste     -9.349534043
```

```
[[1]]
              [,1]
[1,] 9.863232e-01
[2,] 7.175456e-08
[3,] 1.006381e-03
[4,] 1.376937e-02
[5,] 9.900065e-01
[6,] 9.993707e-01
[7,] 9.997900e-01
```

Class Activity 2

Neural Network

Est. Duration: 20 mins

1. Run infert.R
2. We forgot to standardize all the continuous input variables. Since all inputs are positive and not much variation, scale to [0, 1] and then re-run the neural network. What are the differences in the model results before and after standardization.
3. Set a threshold for model prediction to be 1 (e.g. 0.5). Output the confusion matrix. [This has a more direct interpretation than cross entropy].

Instructor answers in infert2.R will be uploaded end of week.

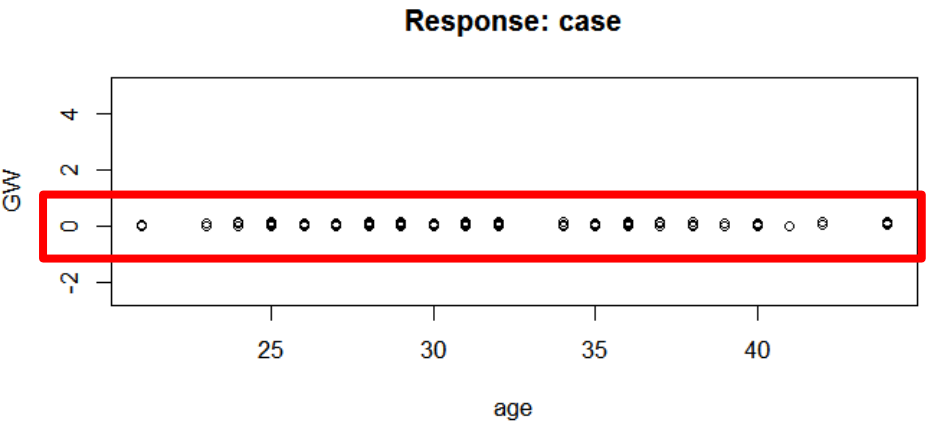
Class Activity 3

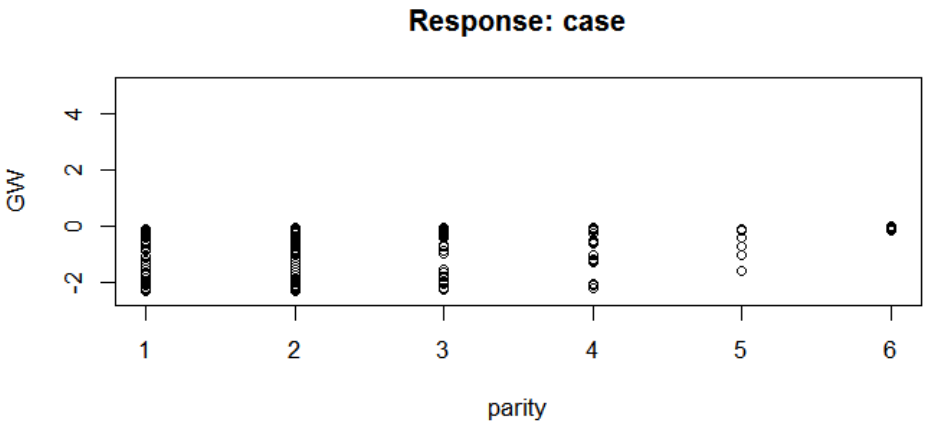
Neural Network

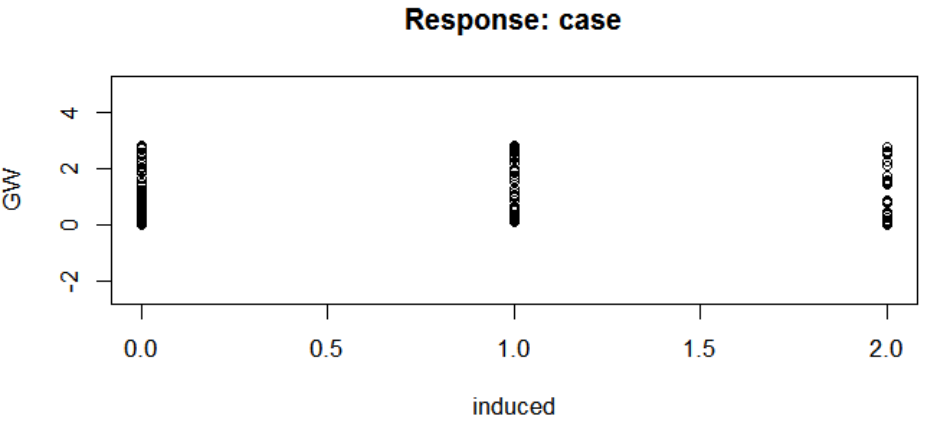
Est. Duration: 15 mins

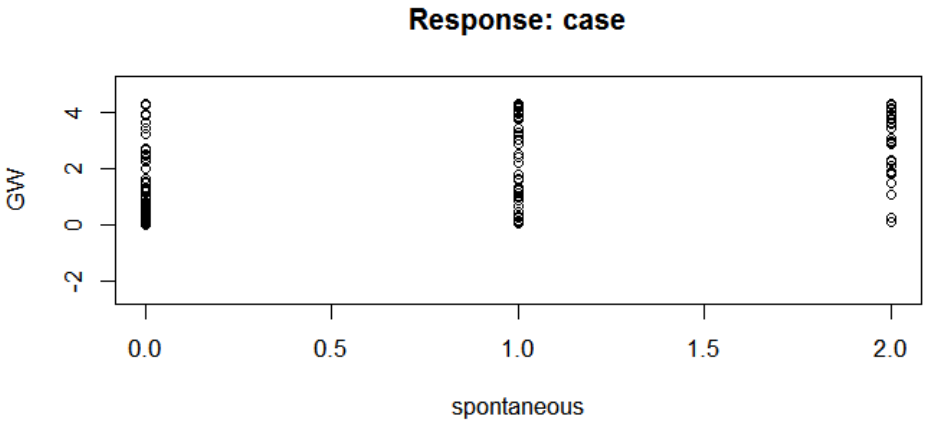
1. Output the infert dataset from R as a CSV file to load into SAS EM.
   - This file is also given as infert.csv
2. Use SAS EM > Model > Neural Network node. What is the result?
   - Try to match the SAS node settings to be as close as neuralnet for comparability

Neuralnet GW results: If no scaling of continuous variables



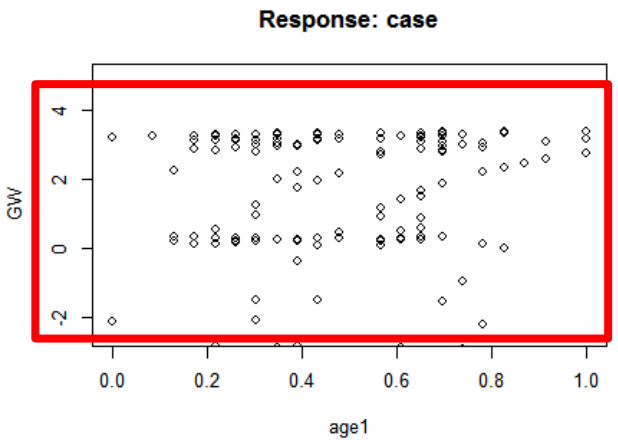


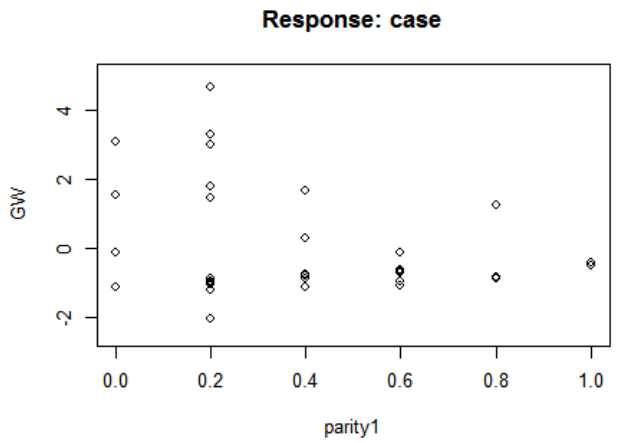


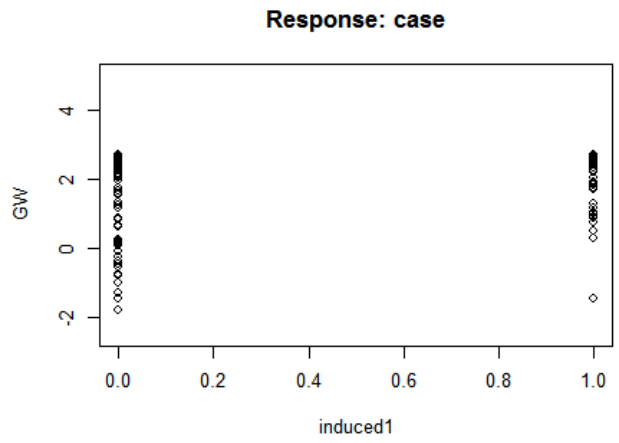


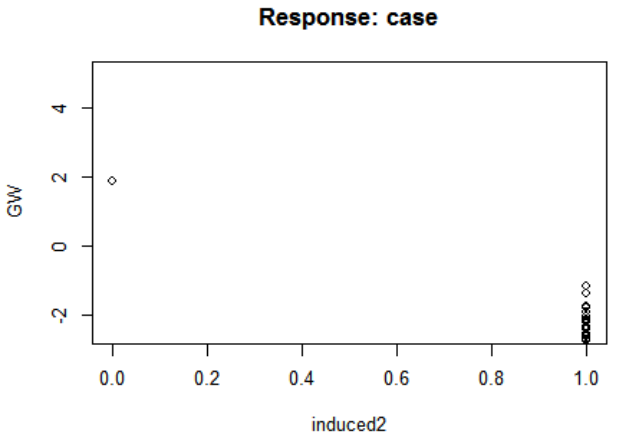
- Without scaling, age is insignificant as all it's values is close to GW = 0
- Source: infert.R

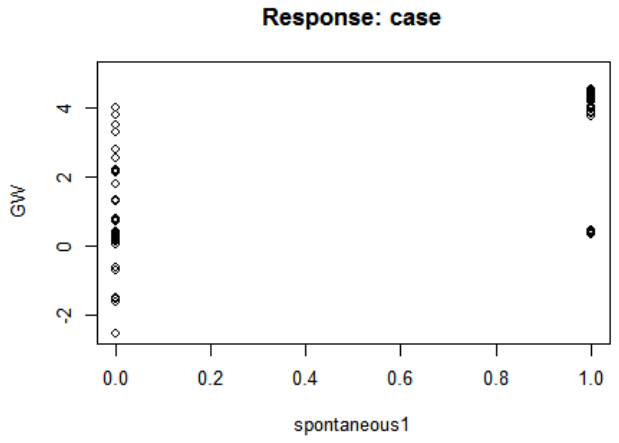
Neuralnet GW results: If continuous variables are scaled

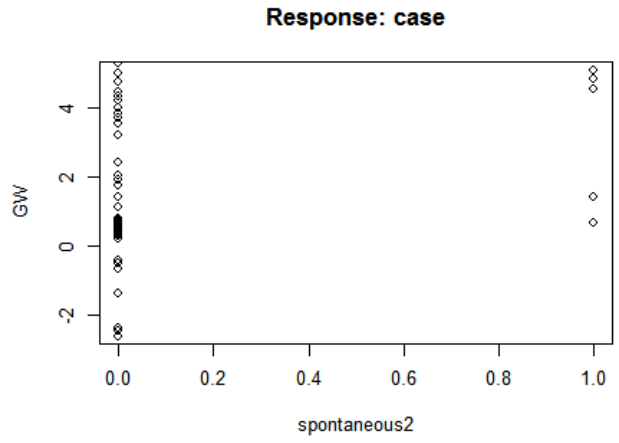






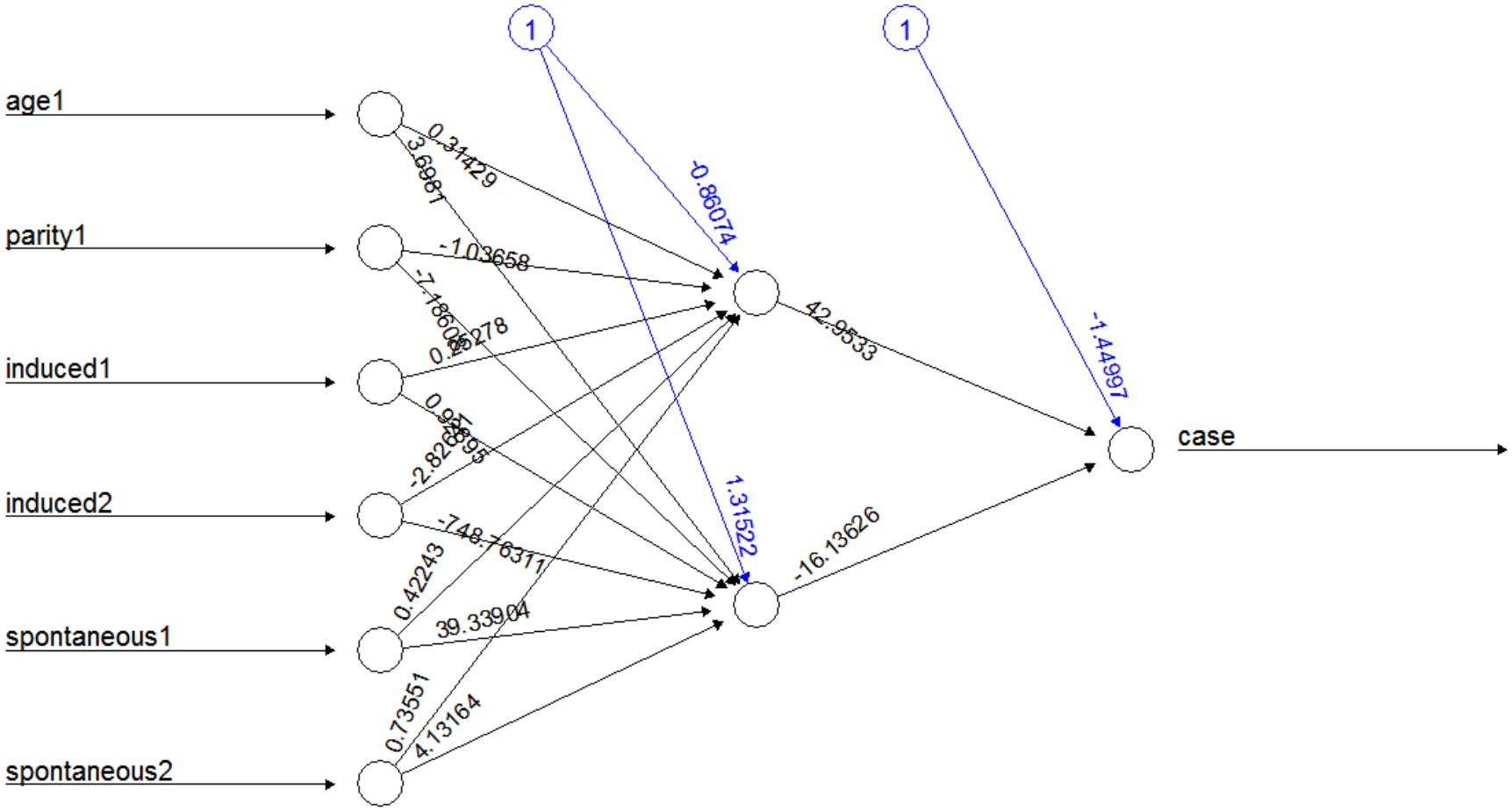






- With scaling, age is significant, and so are parity, induced and spontaneous.
- Source: infert2.R

Neuralnet plot(m2): After continuous variables are scaled and categorical X manually dummied



Error: 116.335566 Steps: 66335

- Source: infert2.R

Comparing Confusion Matrix of Different Neural Networks (on trainset)

R: Unscaled X

R: Scaled X

```
Trainset Confusion Matrix with neuralnet (1 hidden layer,
2 hidden nodes, Unscaled X):
> table(infert$case, pred.m1)
   pred.m1
      0   1
  0 147  18
  1  40  43
> cat('Trainset Confusion Matrix with neuralnet (1 hidden
layer, 2 hidden nodes, Scaled X):')
Trainset Confusion Matrix with neuralnet (1 hidden layer,
2 hidden nodes, Scaled X):
> table(infert$case, pred.m2)
   pred.m2
      0   1
  0 149  16
  1  39  44
```

SAS EM: Scaled X

Event Classification Table

Data Role=TRAIN Target=case Target Label=' '

False Negative	True Negative	False Positive	True Positive
38	147	18	45

- Source: infert2.R, SAS EM Neural Node output