NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Neural Network with R

BC2407 IN-CLASS SESSION SLIDES

Intended Learning Outcomes

Identify aspects of business problems that cause standard analytics models to become useless or less effective.

Apply advanced techniques to overcome or mitigate the weaknesses of standard analytics models.

Evaluate performance of the advanced predictive techniques.

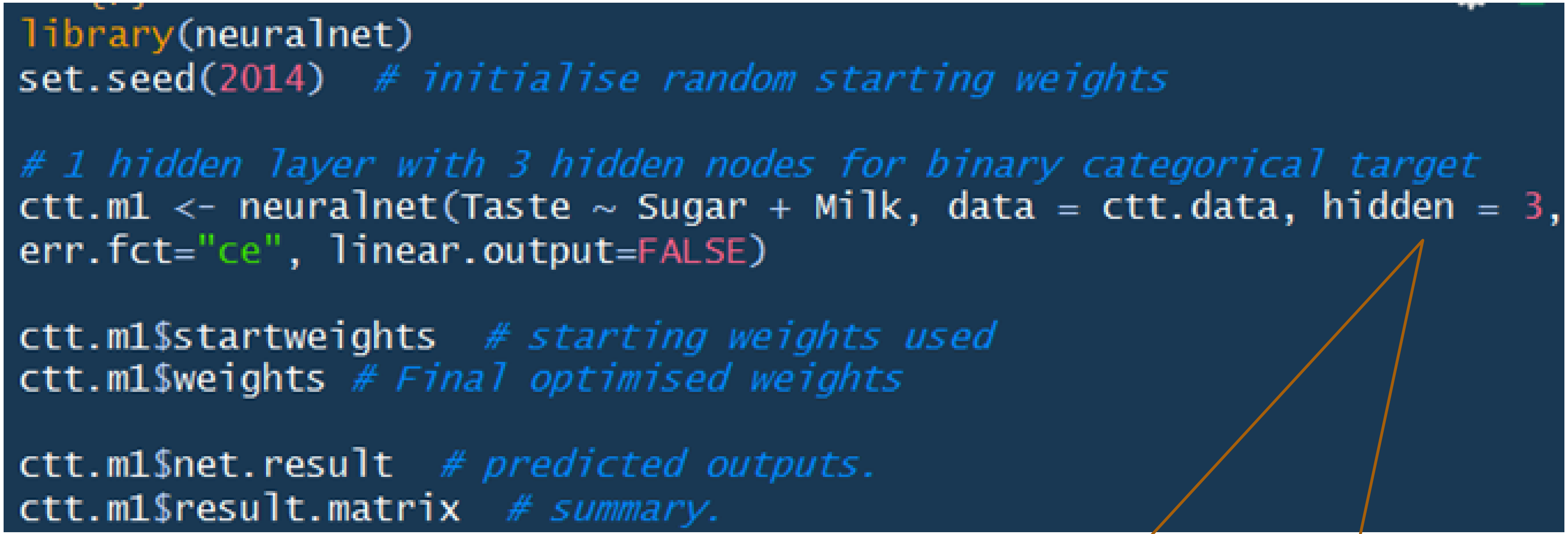
Explain the workings and results of the advanced predictive techniques in the context of the business problem to client/employer.

Propose business solutions/recommendations based on the advanced predictive techniques.

Quiz

Ungraded. Check your understanding of this Session Content.
Use your real name (not nickname) in the quiz.

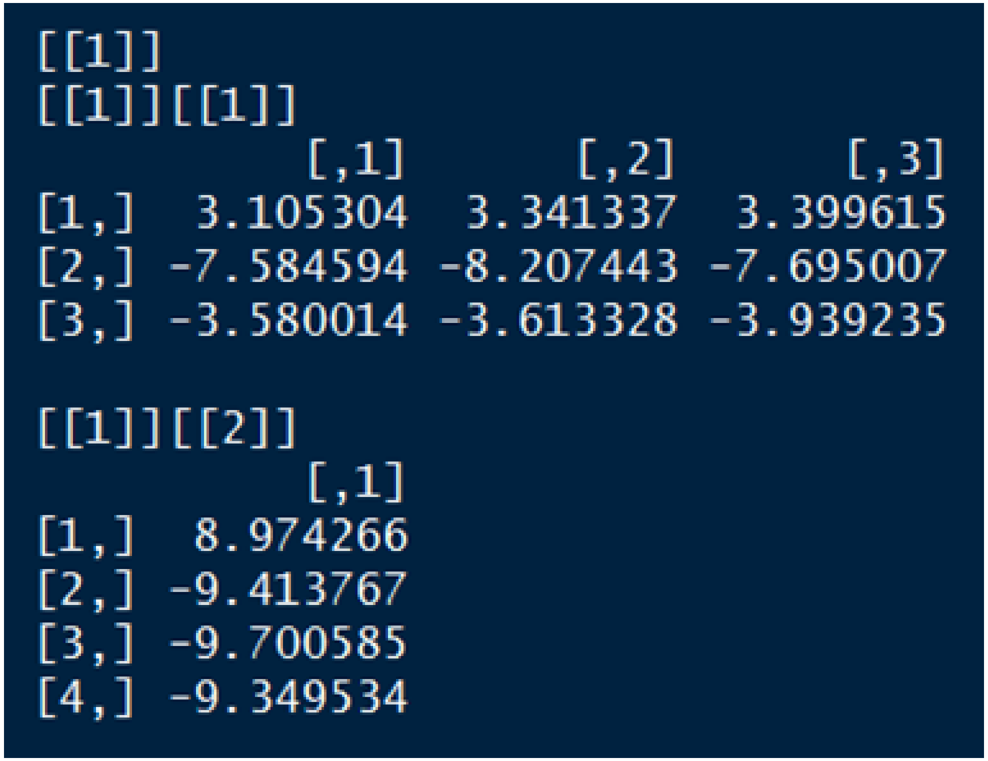
Rpackage neuralnet on chocolate taste test data



```r
library(neuralnet)
set.seed(2014)  # initialise random starting weights

# 1 hidden layer with 3 hidden nodes for binary categorical target
ctt.m1 <- neuralnet(Taste ~ Sugar + Milk, data = ctt.data, hidden = 3,
err.fct="ce", linear.output=FALSE)

ctt.m1$startweights  # starting weights used
ctt.m1$weights # Final optimised weights

ctt.m1$net.result  # predicted outputs.
ctt.m1$result.matrix  # summary.
```

3 hidden nodes in default 1 hidden layer. To specify more than 1 hidden layer, use a vector. E.g. c(4, 2, 2) means 3 hidden layers with 4 nodes in 1st hidden layer, 2 nodes in 2nd hidden layer and 2 nodes in 3rd hidden layer.

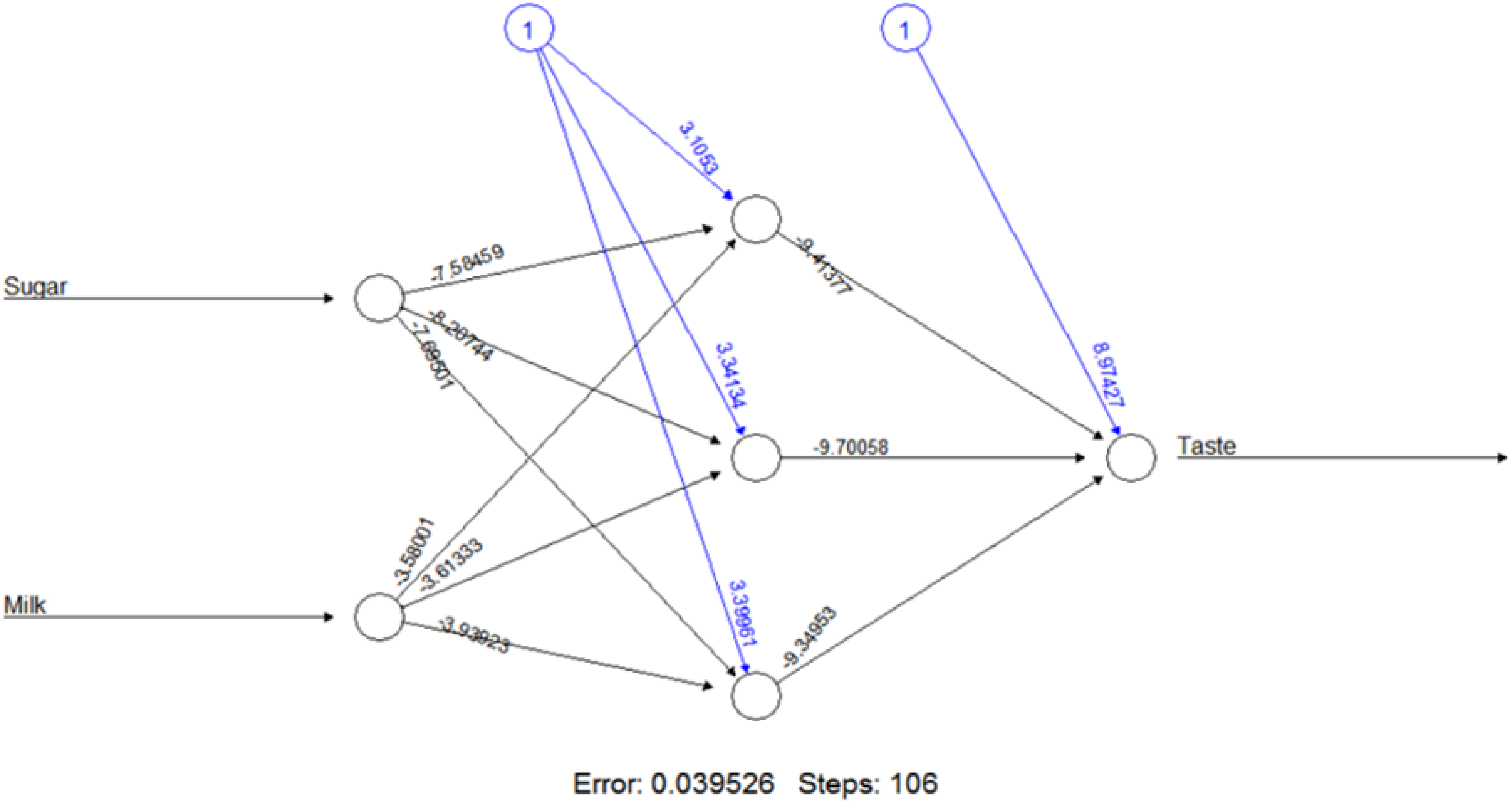
ctt.m1$weights shows the list of final optimized weights



```
[[1]]
[[1]][[1]]
          [,1]      [,2]      [,3]
[1,]  3.105304  3.341337  3.399615
[2,] -7.584594 -8.207443 -7.695007
[3,] -3.580014 -3.613328 -3.939235

[[1]][[2]]
          [,1]
[1,]  8.974266
[2,] -9.413767
[3,] -9.700585
[4,] -9.349534
```

These weights can be visualized on the neural network via the plot() function.

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



Source: Chew C.H. (2020) AI, Analytics and Data Science Vol 1, Chap 9, Figure 9.2.

Q: Using the final weights, write the activation functions in each hidden node and output node?

Class Activity 1

Neural Network Interpretation

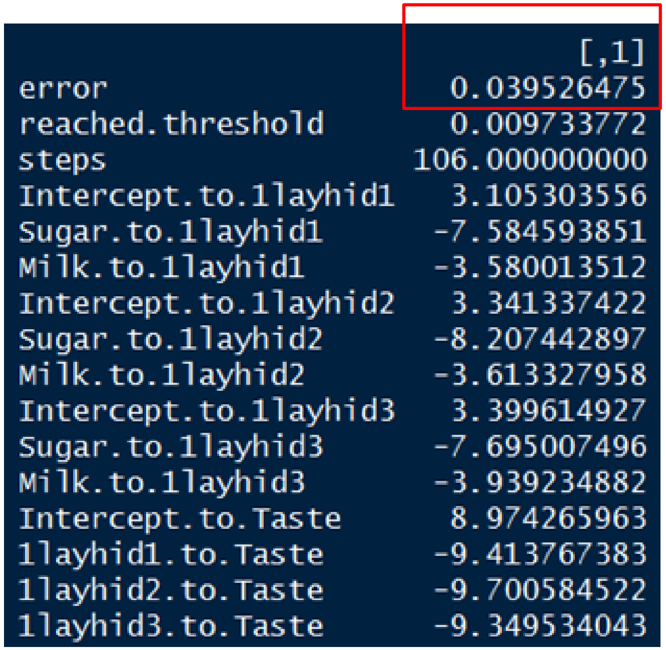
Part of Pre-Class Learning

Est. Duration: 20 mins

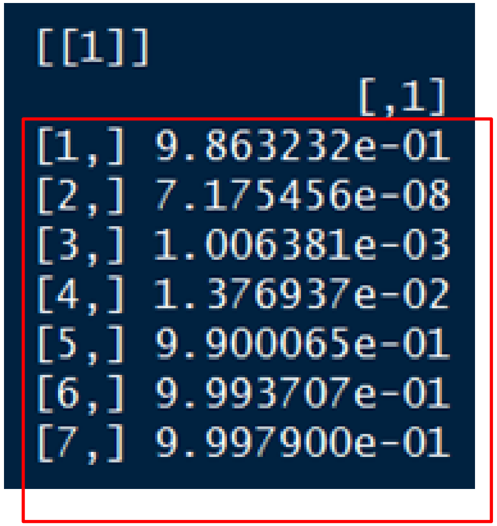
1. The Chocolate Taste Test dataset, randomised start weights and final optimized weights in the neural network above are provided in Excel file CTT.xlsx.

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?

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
```

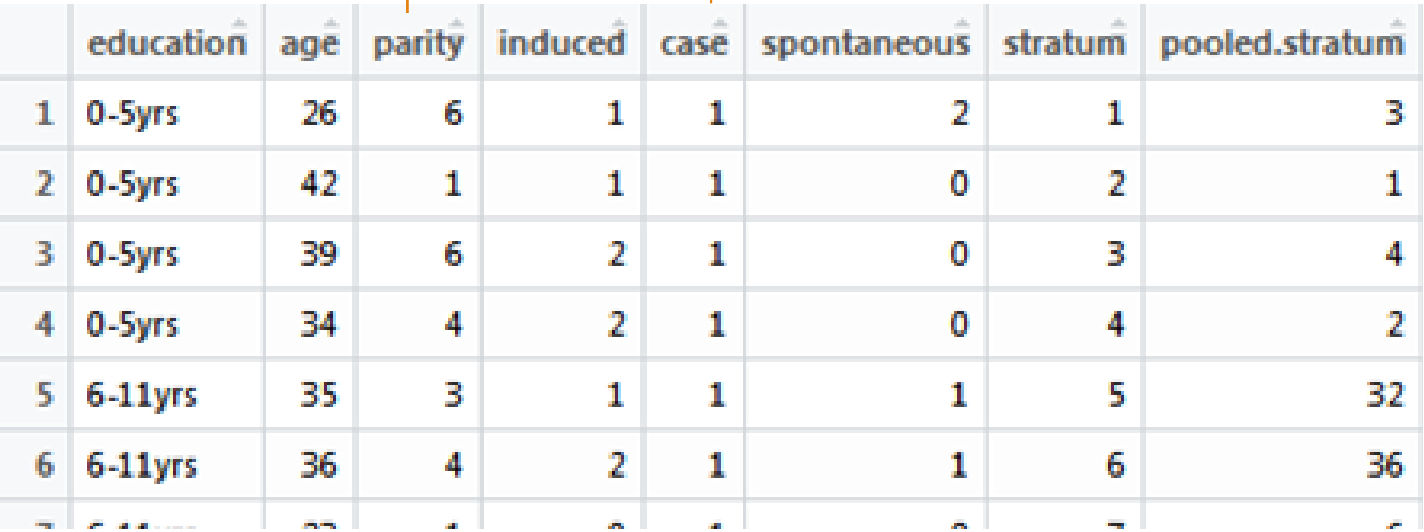
Normalize scale of each input variables

- Weights are critical to neural network performance
  - Sensitive to scale of the input variable X.
  - Example: Age, Annual Income.
- Three popular ways to Scale:
  - [0, 1] Scale. $\frac{X - min(X)}{max(X) - min(X)}$
  - [-1, 1] Scale. $2\frac{X - min(X)}{max(X) - min(X)} - 1$
  - Standard Normal Scale i.e. z score. $\frac{X - mean(X)}{SD(X)}$
    - Use scale() in R.

Activities: Infert dataset in Base R (248 observations)

Num of births

1: Infertile;
0: Not infertile.



	education	age	parity	induced	case	spontaneous	stratum	pooled.stratum
1	0-5yrs	26	6	1	1	2	1	3
2	0-5yrs	42	1	1	1	0	2	1
3	0-5yrs	39	6	2	1	0	3	4
4	0-5yrs	34	4	2	1	0	4	2
5	6-11yrs	35	3	1	1	1	5	32
6	6-11yrs	36	4	2	1	1	6	36

- Source: infert.R
- ?infert to view documentation
- Objective: Predict Infertility risk based on age, births and abortions.

Class Activity 2

Neural Network

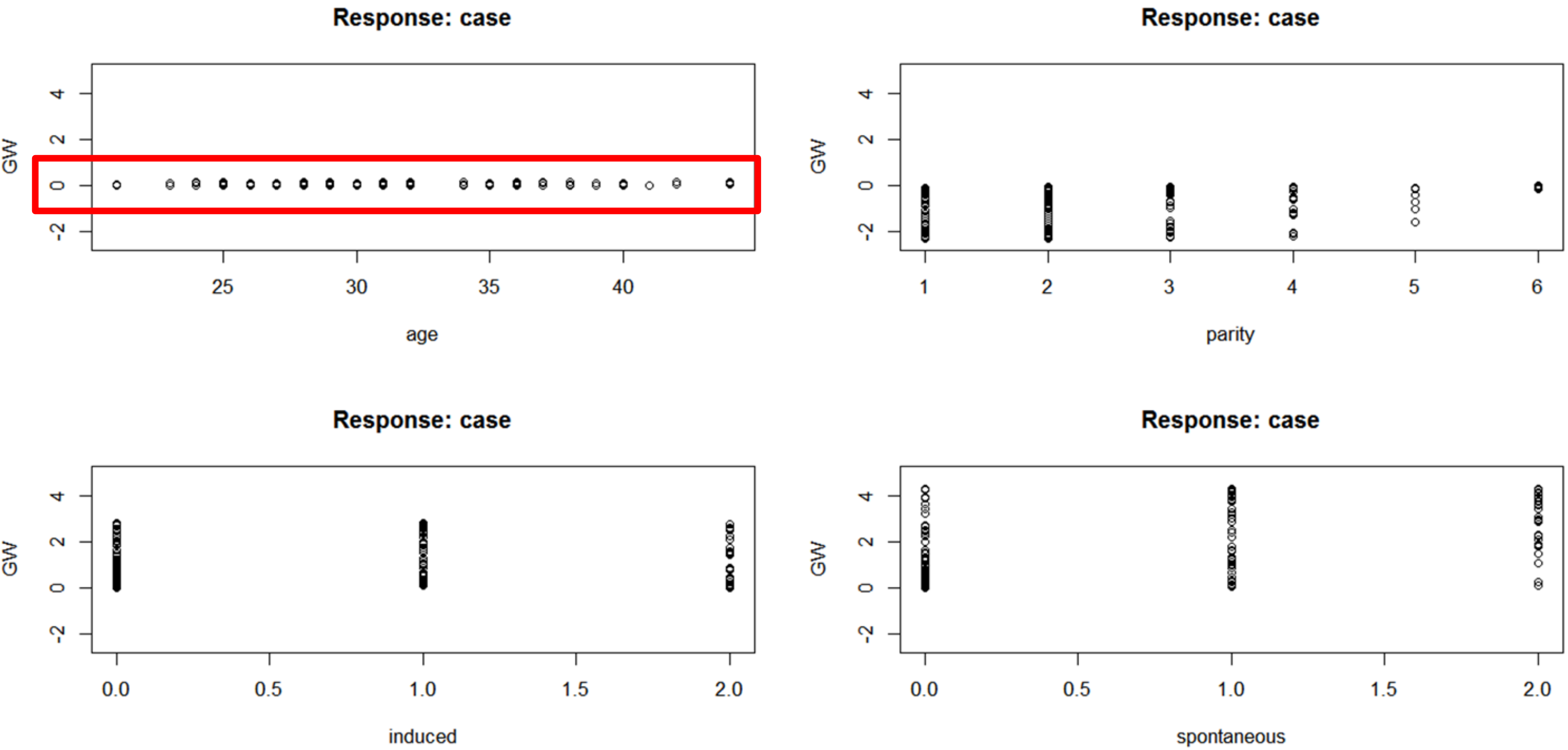
Est. Duration: 20 mins

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.

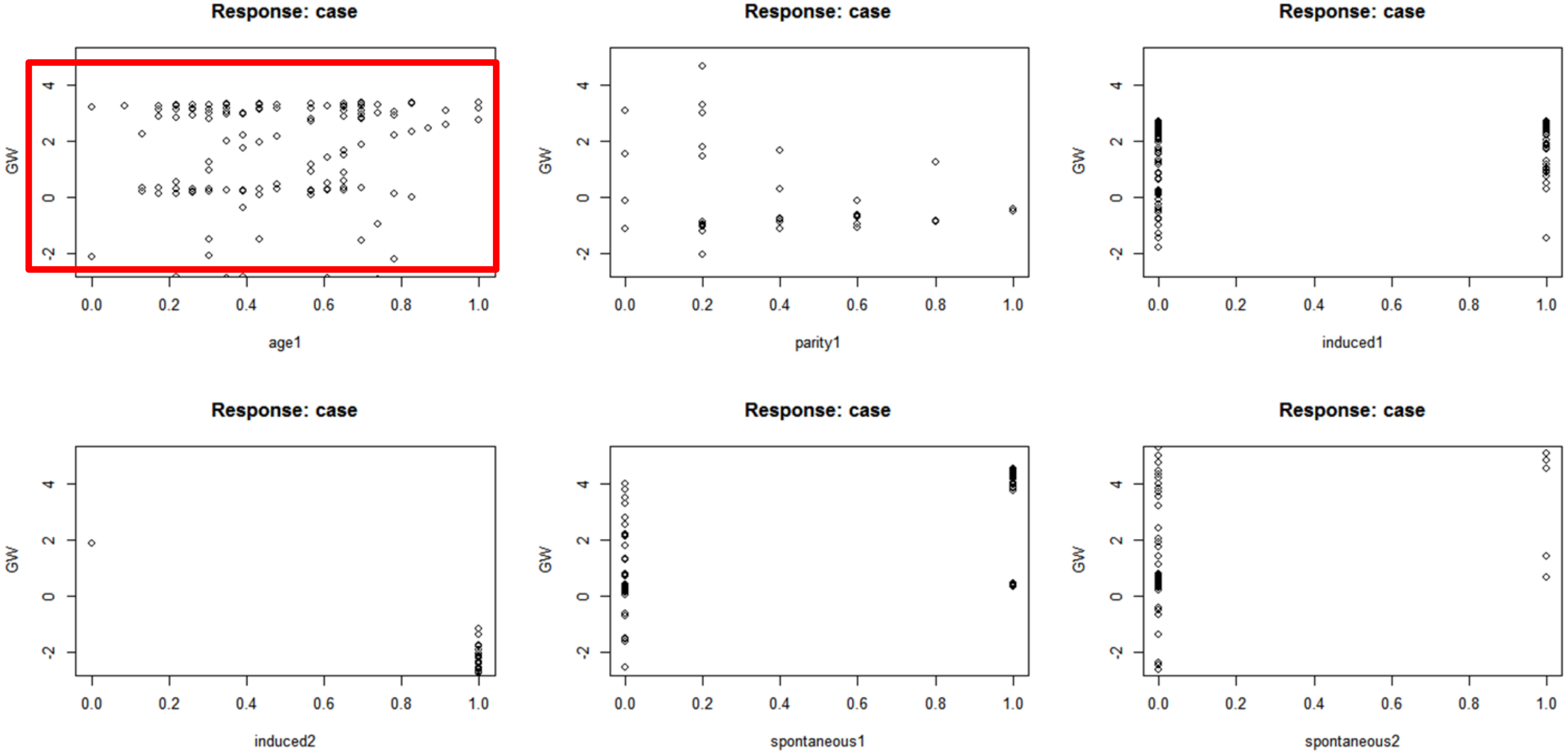
Key Findings of Class Activity 1 & 2

Neuralnet GW results: If no scaling of continuous variables



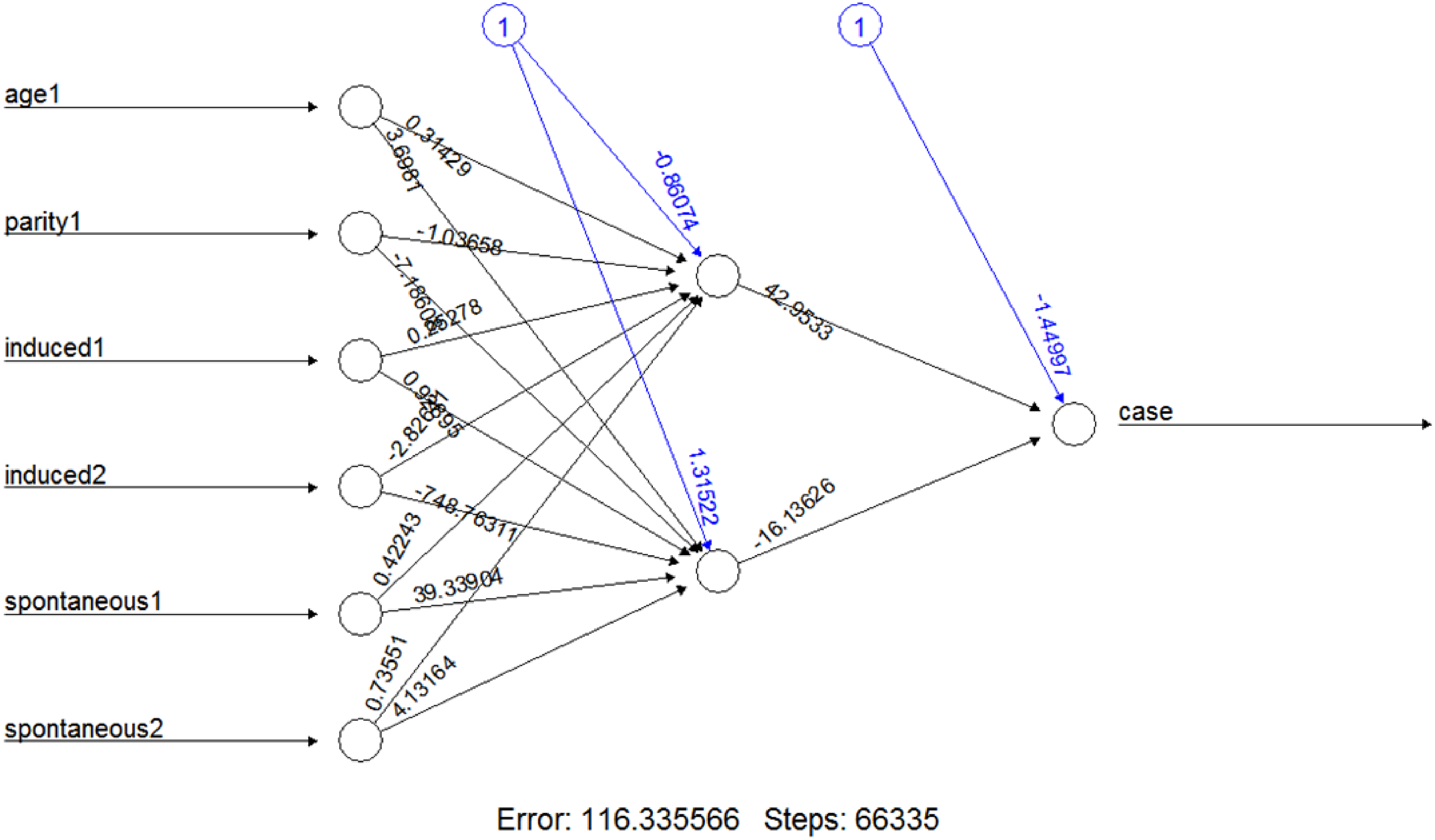
- Without scaling, age is insignificant as all it’s values is close to GW = 0
- Source: infert.R

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

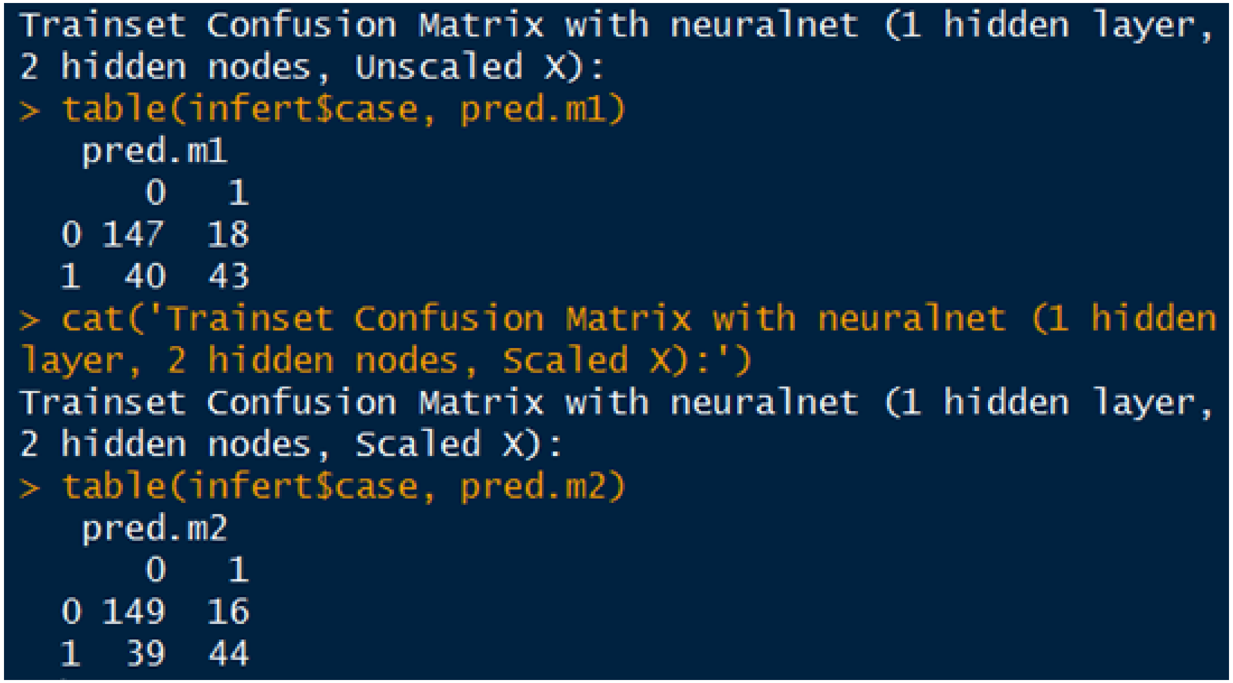


- Source: infert2.R

Comparing Confusion Matrix of Different Neural Networks (on trainset)

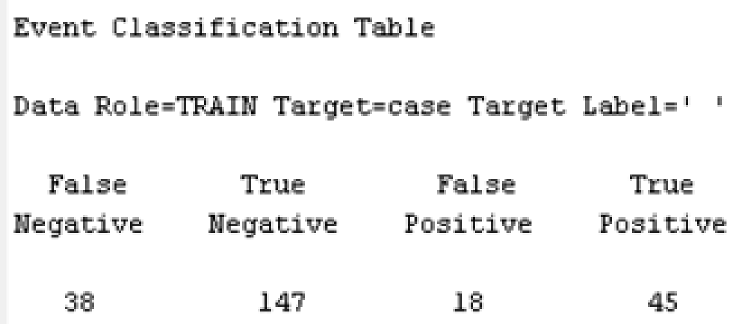
R: Unscaled X

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        True        False        True
Negative     Negative     Positive     Positive

   38          147          18           45
```

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

Summary

- Neural networks can be used for Categorical Y or Continuous Y.
- Can capture a very flexible/complicated relationship between the outcome and a set of predictors.
- The network "learns" and updates its model iteratively as more data are fed into it. Good for sequential learning.
- Major danger: overfitting.
- Good predictive performance, but "black box" in nature.
- R neuralnet can only accept numeric variables. No categorical variables. Thus, need to manually create dummy variables.
- R nnet can accept numeric or categorical variables. But remember to check and set the correct type [e.g. factor(x)]. Else the wrong default algorithm may be applied and you get nonsense results.

Reminder

Please complete the Pre-Class Learning Activities before next class.

Reflection on your Learning

Go

NTULearn Class Site > Journal

Post

Read the instructions and post entry on this week’s learning.

- Reply on the 3 questions as stated in the Journal Instructions.