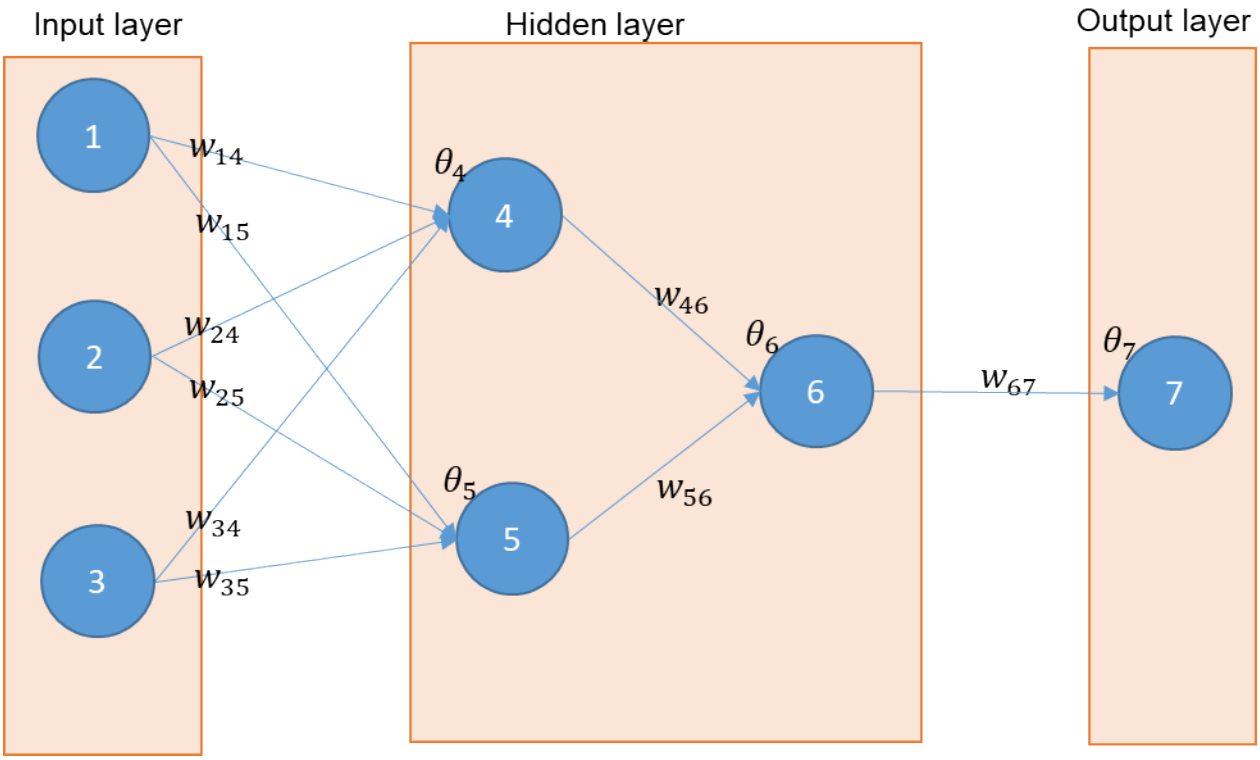
Harshaanth Thiyagaraja Kumar

Please allocate sufficient time for all homework assignments as they usually take time to complete. Start early on assignments. If you have doubts, please email the instructor.

Please submit your answers and required snapshots of reports and charts in MS Word (or PDF) file, together with your Excel workbook.

1. Model estimation.

Describe how to get the weights and bias values in the following model (using equations and bullet points)



The weights q (theta) and w are typically initialized to random values in the range -0.05 to +0.05.

These initial weights are used in the first round of training.

One trial:

- Assign random values to all weights and bias values
- Use the first record to calculate the outputs of the first hidden layer

$$output_4 = \frac{1}{1 + e^{-(\theta_4 + w_{14}x_1 + w_{24}x_2 + w_{34}x_3)}}$$

$$output_5 = \frac{1}{1 + e^{-(\theta_5 + w_{15}x_1 + w_{25}x_2 + w_{35}x_3)}}$$

- Calculate outputs of the second hidden layer

$$output_6 = \frac{1}{1 + e^{-(\theta_6 + w_{46} output_4 + w_{56} output_5)}}$$

- Calculate outputs of the output layer

$$output_7 = \frac{1}{1 + e^{-(\theta_7 + w_{67} output_6)}}$$

- Calculate error

$$err = \hat{y}_k(1 - \hat{y}_k)(y_k - \hat{y}_k) = output_7(1 - output_7)(y - output_7)$$

- Use error to update all weights and bias values

$$\theta_j^{new} = \theta_j^{old} + l * err$$
$$w_{ij}^{new} = w_{ij}^{old} + l * err$$

- Repeat the previous steps and stop following the Common Criteria
- One epoch: repeat for all records
- Stop following the Common Criteria

2. Case study

East-West Airlines has entered into a partnership with the wireless phone company Telcon to sell the latter's service via direct mail. The file EastWestAirlinesNN.xls contains a subset of a data sample of who has already received a test offer. About 13% accepted.

You are asked to develop a model to classify East-West customers as to whether they purchased a wireless phone service contract (target variable Phone_Sale), a model that can be used to predict classifications for additional customers.

2.1 Using XLMiner, fit a logistic regression model (spilt data into 60% training and 40% validation). Use stepwise for variable selection and choose the last model among models recommended by variable selection.

Feature Selection

Best Subsets

Subset ID	Intercept	Topflight	Balance	Qual_miles	cc1_miles?	cc2_miles?	cc3_miles?	Bonus_miles	Bonus_trans	Flight_miles_12mo	Flight_trans_12	Online_12	Email	Club_member	Any_cc_miles_12mo
Subset 1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Subset 2	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0
Subset 3	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1
Subset 4	1	0	0	0	0	0	0	0	1	0	0	1	0	0	1

Best Subsets Details

Subset ID	#Coefficients	RSS	Mallows's Cp	Probability
Subset 1	1	3098.0758	84.5961253	2.5007E-14
Subset 2	2	3028.3825	17.45350578	0.008031268
Subset 3	3	3016.1521	7.319726149	0.177772121
Subset 4	4	3011.3904	4.595594709	0.395193123

2.1.1 Which variables are included in the **selected model**

Bonus_trans, Online_12, Any_cc_miles_12mo

2.1.2 Report the selected model's regression model (the equation and the coefficient table)

Coefficients

Predictor	Estimate	Confidence Interval: Lower	Confidence Interval: Upper	Odds	Standard Error	Chi2-Statistic	P-Value
Intercept	-2.506199827	-2.708078678	-2.304320976	0.08157766	0.103001306	592.0330156	9.1E-131
Bonus_trans	0.029615335	0.017933647	0.041297024	1.030058231	0.005960155	24.68983815	6.73E-07
Online_12	0.160065349	0.018710211	0.301420487	1.173587561	0.072121294	4.925696333	0.02646
Any_cc_mile	0.497474451	0.232882467	0.762066434	1.644562598	0.134998391	13.57951879	0.000229

Logit(Phone_sale=1) = -2.5062 + 0.0296*Bonus_trans + 0.16*Online_12 + 0.4974*Any_cc_miles_12mo

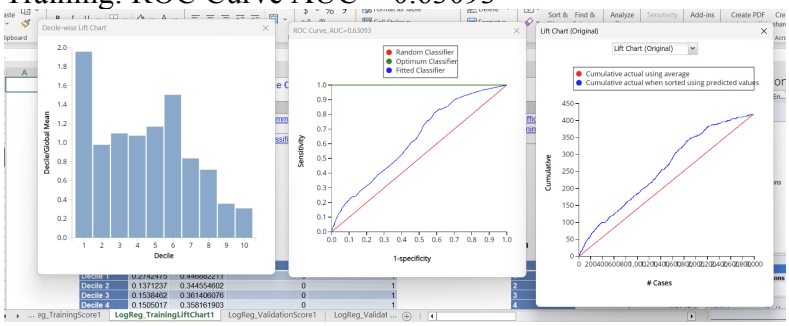
Odds(Phone_sale=1) = $e^{-2.5062 + 0.0296*Bonus\_trans + 0.16*Online\_12 + 0.4974*Any\_cc\_miles\_12mo}$

Probability (Phone_sale=1) =

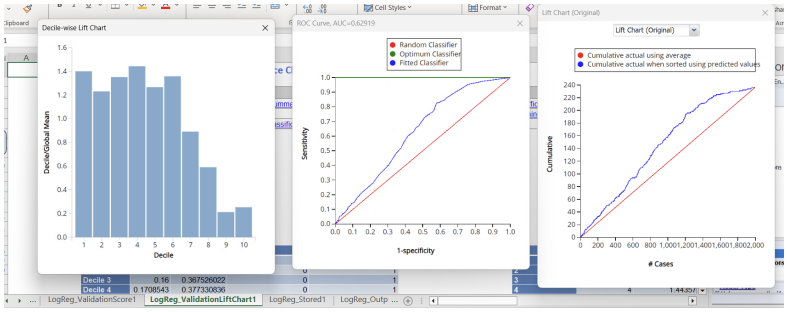
$$\frac{1}{1+e^{-(-2.5062 + 0.0296*\text{Bonus\_trans} + 0.16*\text{Online\_12} + 0.4974*\text{Any\_cc\_miles\_12mo})}}$$

2.1.3 Report the validation lift charts. How does this model work?

Training: ROC Curve AUC = 0.63093



Validation: ROC Curve AUC = 0.62919



The ROC curve indicates that the model is a decently reliable predictor of Phone_Sale. Both the ROC curve for the training data (AUC = 0.63093) and the validation data (AUC = 0.62919) exhibit close similarity, suggesting no major overfitting concerns. This indicates the model performs effectively on both the training and validation data, indicating its robustness with new data.

A stepper slope in the lift chart throughout most of the decile range means that the logistic regression model is good at identifying the decile groups that are most likely to contain the desired outcome.

2.1.4 What are the overall accuracy and sensitivity of the validation dataset? Explain in your own words, what does sensitivity mean in this example? If the company wants to pick out potential customers who would purchase the phone contract, does this model work well? How to change the cutoff value to serve the purpose of the firm (up or down)?

Validation: Classification Summary

Confusion Matrix

Actual\Predicted	0	1
0	1757	1
1	235	1

Error Report

Class	# Cases	# Errors	% Error
0	1758	1	0.056882821
1	236	235	99.57627119
Overall	1994	236	11.83550652

Metrics

Metric	Value
Accuracy (#correct)	1758
Accuracy (%correct)	88.16449348
Specificity	0.999431172
Sensitivity (Recall)	0.004237288
Precision	0.5
F1 score	0.008403361
Success Class	1
Success Probability	0.5

Overall accuracy = (TP+TN) / (TP+TN+FP+FN) = (1+1757) / (1+1757+1+235) = 88.16%

Sensitivity (Recall) = TP / (TP+FN) = 1 / (1+235) = 0.42%

Even though the overall accuracy of the model in predicting correct outcomes is 88.16%, the sensitivity (recall), representing the model's ability to detect true positive cases (customers who would purchase the phone contract) among all actual positive cases, is just 0.42%.

This low sensitivity occurs because the model has predominantly classified over 99% of the cases as classification 0, either true negatives or false negatives. While this model effectively predicts customers who would not purchase the phone contract, it falls short in identifying those who would.

To improve the prediction of classification 1 and boost Sensitivity (Recall), we need to lower the cutoff value. This adjustment aligns with the firm's objective of pinpointing potential customers likely to purchase the phone contract.

2.1.5 Choose the cutoff values based on lift charts and report the new overall accuracy and sensitivity.

Decile Chart Data

ID	Decile	Decile/Global Mean
1	1	1.401115748
2	2	1.231283536
3	3	1.351864407
4	4	1.443573801
5	5	1.267372881
6	6	1.358657695
7	7	0.891619113
8	8	0.591440678
9	9	0.212290265
10	10	0.253474576

Step 1: Based on Lift chart, Deciles 1-6 have a Global mean > 1, so Sorting prob of class 1 from largest to smallest and choosing the record at 60th % and looking at its prob of class 1

Step 2: New probability cutoff = 0.12205

Validation: Classification Summary

Confusion Matrix		
Actual\Predicted	0	1
0	837	921
1	61	175

Error Report			
Class	# Cases	# Errors	% Error
0	1758	921	52.3890785
1	236	61	25.84745763
Overall	1994	982	49.24774323

Metrics	
Metric	Value
Accuracy (#correct)	1012
Accuracy (%correct)	50.75225677
Specificity	0.476109215
Sensitivity (Recall)	0.741525424
Precision	0.159671533
F1 score	0.262762763
Success Class	1
Success Probability	0.12205

Overall accuracy =(TP+TN) / (TP+TN+FP+FN) = (175+837) / (175+837+921+61) = 50.75%

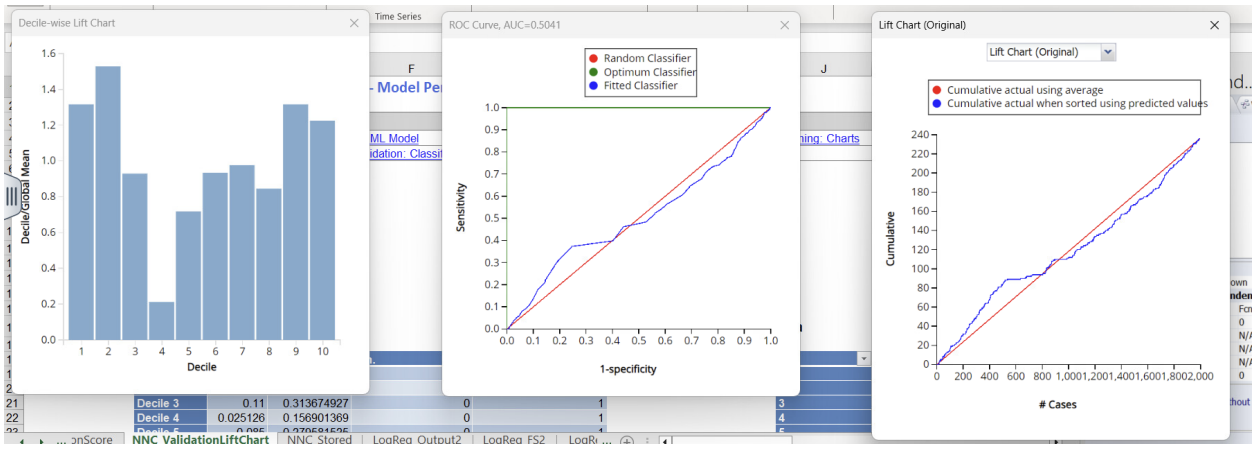
Sensitivity (Recall) = TP / (TP+FN) = 175 / (175+61) = 74.15%

Even though the overall accuracy of the model in predicting correct outcomes has reduced to 50.75%, the sensitivity (recall), representing the model's ability to detect true positive cases (customers who would purchase the phone contract) among all actual positive cases, has significantly increased to 74.15%. The model is identifying 74.15% of customers who would purchase the phone contract.

This higher sensitivity is because we reduced the cutoff to 0.12205 the model has classified more of the cases as classification 1, hence making this model more aligned with the firm's objective of pinpointing potential customers likely to purchase the phone contract.

2.2 Using XLMiner, run a neural net model on the **partitioned** data, using the same set of variables in **2.1.1**, using the option to standardize the data, setting one hidden layer, 5 nodes in the hidden layer, 100 epochs. Report the following:

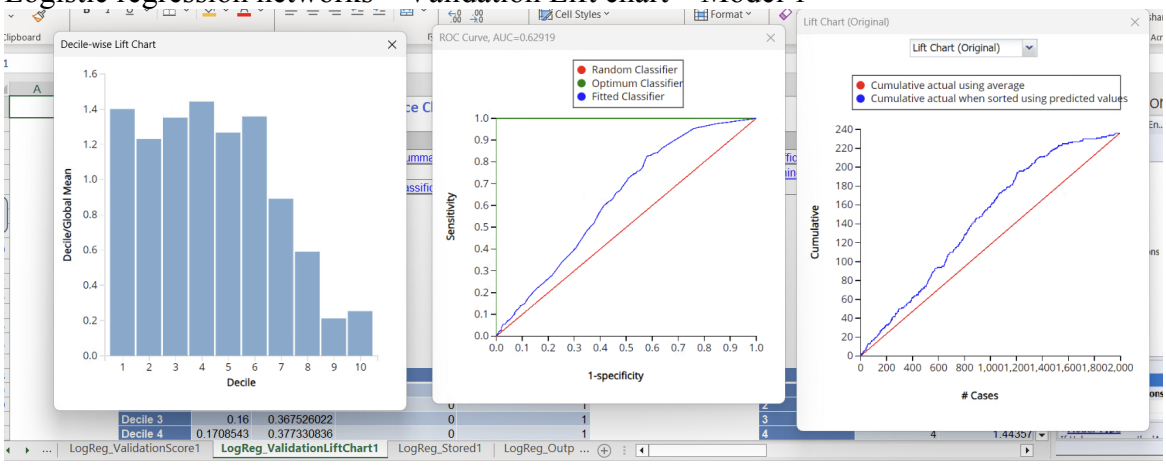
2.2.1 Report the validation lift charts. How does this model work?



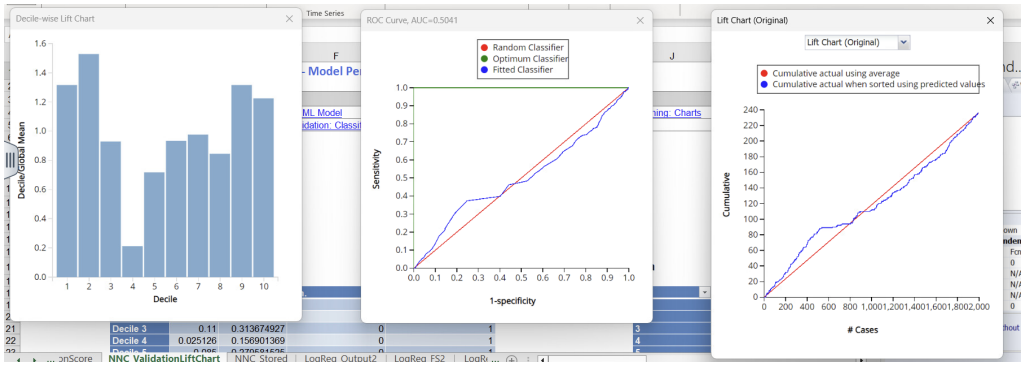
The model appears to perform poorly. The Area Under the Curve (AUC) of 0.5041 for a Receiver Operating Characteristic (ROC) curve indicates that the model performs no better than random guessing. It seems unlikely that the model can effectively predict phone sales.

2.2.2 Compare the lift charts using neural networks and logistic regression, which model works better?

Logistic regression networks – Validation Lift chart – Model 1



Neural networks – Validation Lift chart – Model 2



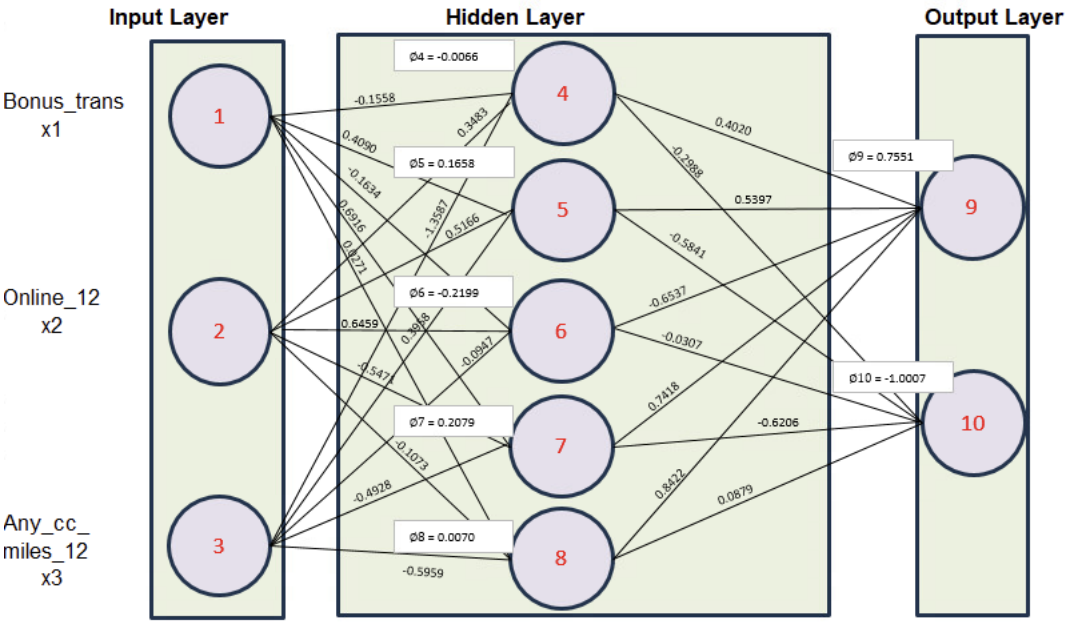
The lift chart for the logistic regression model (Model 1) shows a steeper slope than the lift chart for the neural network model (Model 2) throughout most of the decile range. This means that the logistic regression model is better at identifying the decile groups that are most likely to contain the desired outcome.

2.2.3 Draw a neural network for this problem (You can hand draw the diagram or use any drawing software).

Neuron Weights

Neuron Weights: Input Layer - Hidden Layer 1				
Neurons	Bonus_trans	Online_12	Any_cc_miles_12mo	Bias
Neuron 1	-0.155825674	0.34837156	-1.358798983	-0.0066465
Neuron 2	0.409044248	0.51660389	0.395837765	0.1658085
Neuron 3	-0.163495386	0.64591553	-0.094753301	-0.2199716
Neuron 4	0.691698545	-0.5471572	-0.492846584	0.20793564
Neuron 5	0.027159676	-0.107347	-0.595905101	0.00708039

Neuron Weights: Hidden Layer 1 - Output Layer						
Neurons	Neuron 1	Neuron 2	Neuron 3	Neuron 4	Neuron 5	Bias
0	0.402065445	0.53976899	-0.653768969	0.74186219	0.84221047	0.755109
1	-0.298851876	-0.5841963	-0.030758085	-0.6206669	0.08793622	-1.00075



2.2.4 If a customer had 10 non-flight bonus transactions in the past 12 months, had 1 online purchase within the past 12 months, didn't add miles on any credit card type within the past 12 months, would he purchase Telcom service as a result of the direct mail campaign?

Input Layer:
X1 = 10
X2 = 1
X3 = 0

Hidden Layer:

$$output_4 = \frac{1}{1 + e^{-(\theta_4 + w_{14}x_1 + w_{24}x_2 + w_{34}x_3)}}$$

$$output_5 = \frac{1}{1+e^{-(\theta_5 + w_{15}x_1 + w_{25}x_2 + w_{35}x_3)}}$$

$$output_6 = \frac{1}{1+e^{-(\theta_6 + w_{16}x_1 + w_{26}x_2 + w_{36}x_3)}}$$

$$output_7 = \frac{1}{1+e^{-(\theta_7 + w_{17}x_1 + w_{27}x_2 + w_{37}x_3)}}$$

$$output_8 = \frac{1}{1+e^{-(\theta_8 + w_{18}x_1 + w_{28}x_2 + w_{38}x_3)}}$$

$$output_4 = \frac{1}{1+e^{-(-0.006+(-0.1558)10+(0.3483)1+(-1.3587)0)}} = 0.2286$$

$$output_5 = \frac{1}{1+e^{-(0.1658+(0.4090)10+(0.5166)1+(0.3958)0)}} = 0.9911$$

$$output_6 = \frac{1}{1+e^{-(-0.2199+(-0.1634)10+(0.6459)1+(-0.0947)0)}} = 0.1496$$

$$output_7 = \frac{1}{1+e^{-(0.2079+(0.6916)10+(-0.5471)1+(-0.4928)0)}} = 0.0021$$

$$output_8 = \frac{1}{1+e^{-(0.0070+(0.0271)10+(-0.1073)1+(-0.5959)0)}} = 0.4605$$

Output Layer:

$$output_9 = \frac{1}{1 + e^{-(\theta_9 + w_{49}output_4 + w_{59}output_5 + w_{69}output_6 + w_{79}output_7 + w_{89}output_8)}}$$

$$output_{10} = \frac{1}{1 + e^{-(\theta_{10} + w_{410}output_4 + w_{510}output_5 + w_{610}output_6 + w_{710}output_7 + w_{810}output_8)}}$$

$$output_9 = \frac{1}{1+e^{-(0.7551+(0.4020)0.2286+(0.5397)0.9911+(-0.6537)0.1496+(0.7418)0.0021+(0.8422)0.4605)}} = 0.5812$$

$$output_{10} = \frac{1}{1+e^{-(-1.0007+(-0.2988)0.2286+(-0.5841)0.9911+(-0.0307)0.1496+(-0.6206)0.0021+(0.0879)0.4605)}} = 0.1662$$

Output 9 > 0.5 > Output 10

Therefore Classification 0,
he wouldn't purchase Telcom service as a result of the direct mail campaign.

Submission checklist

[ ] Word or PDF file with answers to questions 1-6, including snapshots of required reports charts.
[ ] Excel Workbooks.