



THE LONDON SCHOOL  
OF ECONOMICS AND  
POLITICAL SCIENCE ■

---

**MODULE 7 UNIT 2**  
**Activity submission**

**Learning outcomes:**

**LO2:** Recognise the model mechanics of neural networks.

**LO3:** Apply R code to train a neural network to make a prediction.

**LO4:** Interpret the predictions of a neural network model to inform business decisions.

**Plagiarism declaration**

1. I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own.

2. This assignment is my own work.

3. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as their own work.

4. I acknowledge that copying someone else's assignment (or part of it) is wrong and declare that my assignments are my own work.

**Name: Vasileios Tsoumpris**

## 1. Instructions and guidelines (Read carefully)

### Instructions

1. Insert your name and surname in the space provided above, as well as in the **file name**. Save the file as: **First name Surname M7 U2 Activity Submission – e.g. Lilly Smith M7 U2 Activity Submission**. **NB:** Please ensure that you use the name that appears in your student profile on the Online Campus.
2. Write all your answers in this document. There is an instruction that says, "Start writing here" under each question. Please type your answer there.
3. Submit your assignment in **Microsoft Word only**. No other file types will be accepted.
4. Do **not delete the plagiarism declaration** or the **assignment instructions and guidelines**. They must remain in your assignment when you submit.

**PLEASE NOTE:** Plagiarism cases will be investigated in line with the Terms and Conditions for Students.

**IMPORTANT NOTICE:** Please ensure that you have checked your course calendar for the due date for this assignment.

## Guidelines

1. There are seven pages and two questions in this assignment.
2. Make sure that you have carefully read and fully understood the questions before answering them. Answer the questions fully but concisely and as directly as possible. Follow all specific instructions for individual questions (e.g. “list”, “in point form”).
3. Answer all questions in your own words. Do not copy any text from the notes, readings, or other sources. **The assignment must be your own work only.**
4. At the end of your assignment, please provide feedback on areas where you require further assistance or would like the Assessor to expand on.

## 2. Mark allocation

Each question receives a mark allocation. However, you will only receive a final percentage mark and will not be given individual marks for each question. The mark allocation is there to show you the weighting and length of each question.

Question 1	24
Question 2	24
<b>TOTAL</b>	<b>48</b>

## 3. Questions

In Module 3, you explored the viability of including life insurance in your organisation’s product offering. You used the regression model to calculate the amount that your company would have to pay out if it offered life insurance to its existing customer base.

In the assessment IDE activity, you fitted the regression model and the neural network onto the term life data set. You can now choose the more appropriate model for the task.

### Question 1

Analyse the output of the IDE activity where the error rate of both models were compared, and answer the following questions.

- 1.1 Are there any advantages associated with fitting a neural network on the term life data set compared to other models? If so, discuss at least one advantage of neural networks compared to other models. If not, substantiate your answer.

(Max. 200 words)

Start writing here:

As depicted in the dataset we are working on, one of the many advantages of neural networks is that they can easily understand the complex relationships of the variables in a dynamic way. In particular, we have seventeen features and one response on the term life insurance dataset. We deal with so many input variables that other models, such as linear

regression, would have difficulty handling. For example, the education of the respondent's spouse or the number of household members are variables that a linear regression could not model. This is because their relationships to the face variable are not linear.

The combination of the bias threshold in the neural networks and their activation functions allows them to generalise well and not overfit in the data. In this way, the neural networks do the non-linear transformation to the input variables and can learn much more complex tasks than other machine learning models, achieving excellent training performance and great accuracy.

The structure of the neural networks (the interconnected layers which contain the perceptrons) gives them the advantage to map and, in extension, model the complex structure of input variables to predict one output, in this case, the face variable.

1.2 What are the disadvantages associated with fitting a neural network on the term life data set compared to other models? Discuss at least two disadvantages and substantiate your answer with specific reference to the term life data set. The disadvantages do not need to be limited to the IDE activity.

(Max. 200 words)

Start writing here:

To start with, a neural network is a black box model which is hard to interpret. Neural networks are less explainable than other models like linear regression. Given the coefficients that linear regression provides in the summary of its output, one can understand how linear regression maps the variables and how the result is generated. On the other hand, when we train a neural network model, we cannot validate how the perceptrons in the hidden layers create their outputs and, by extension, how the final outcome is generated. The black-box nature of neural networks is one of the most significant disadvantages of neural networks.

Moreover, neural networks do need many data to provide good performance. When neural networks are trained in small datasets like that we worked on (only 500 households), they create biased outputs resulting in significant error rates.

Furthermore, neural networks are more computationally expensive than other ML methods. As a result, they are slower to train, especially if they have a lot of hidden layers or a considerable number of neuron nodes.

Lastly, neural networks have numerous parameters to tune, leading to a more comprehensive development process than other simpler algorithms that could provide similar results.

## Question 2

In Module 3, you used a regression model to calculate the policy face value for certain individuals. These individuals, and their predicted outputs, are listed in the following table.

Observation	Income	Education	Numhh	Predicted output
1	US\$100,000	16 years	5	US\$488,419
2	US\$110,000	18 years	6	US\$1,050,567
3	US\$90,000	15 years	3	US\$204,520

Refer to the RMSE rate and the predictions calculated by the neural network for these observations in the IDE activity, and answer the following questions.

- 2.1. Which model do you think is the most appropriate to calculate predictions and inform your decision of expanding your organisation's current product offerings? Substantiate your answer.

(Max. 200 words)

Start writing here:

In my opinion, linear regression is the most appropriate model to use in a business context to help the organisation make decisions about its future strategy. The exploratory data analysis process in the insurance dataset depicts some insights into understanding the complexity of the data and the problem in general. In this case, the business problem seems simple and linear. As a result, the linear regression model gives good results with only three features selected from the dataset.

Therefore, there is no need to choose a complex model, such as a neural network. Neural networks have a much slower development process (a lot of tuning in its parameters), a black-box nature (low explainability) and a high computational cost derived from their hidden layers and the perceptrons, which try to understand non-linear structures of the data variables. Thus, they could be used for more difficult tasks like image classification, where there is a plethora of complex data.

Concluding, the simplicity and interpretability that linear regression provides, in combination with its satisfactory performance and fast predictions, is what an organisation needs to form and design its strategic expansion of the current product offerings.

- 2.2. Compare the **face** values calculated by the neural network and the regression model in Module 3. Which model's predictions would you use to make a decision? Substantiate your answer.

(Max. 200 words)

Start writing here:

In the life insurance dataset, the neural network fits better with the training data than linear regression. Still, it does not necessarily give better results in prediction. Compared with linear regression, due to the small amount of data we input into the neural network, the model seems biased and does not respond well when we make new predictions from predData. This is contradictory with the better RMSE that the model gives compared to linear regression (implemented in module 7).

Furthermore, the Linear regression model (of module 3) is quite sensitive when changing its three specific variables (Numhh, Income, Education) as this model was created based on them. These variables were processed accordingly to overcome problems that could arise when inserted into the model (i.e. skewed data).

On the other hand, the NN seems to map different and more complex relationships between the data. As a result, the NN doesn't create different predictions by changing the Numhh, Income and Education variables while leaving all the others constant. In other words, to create different predictions with a NN, one should tweak many other input variables.

Concluding, I would make a decision based on the linear regression model's predictions. Linear regression is much simpler, interpretable and well-generalised in this specific dataset.

## 4. Rubric

The following rubric will be used to grade your submission for this activity submission.

	Unsatisfactory	Limited	Accomplished	Exceptional
<b>Question 1.1:</b> <b>Adherence to the brief</b> <i>The answer successfully compares the effectiveness of the regression model and the neural network.</i>	No submission. OR Answer fails to adhere to any of the elements contained in the brief. (0)	Answer adheres to some elements contained in the brief, but some key elements are missing. Answer does not fall within the prescribed word count (50 words over the word count). (2)	Answer adheres to most, but not all, elements of the brief. Almost all information is provided and relevant. Answer falls within the prescribed word count. (4)	Answer adheres to all elements of the brief. All information provided is comprehensive and relevant. Answer falls within the prescribed word count. (6)
<b>Question 1.2:</b> <b>Adherence to the brief</b>	No submission. OR	Answer adheres to some elements contained in the brief, but some key	Answer adheres to most, but not all, elements of the brief. Almost all	Answer adheres to all elements of the brief. All information provided is

<p><i>The answer elaborates on the advantages of the neural network over other models, or substantiates why there are no advantages.</i></p>	<p>Answer fails to adhere to any of the elements contained in the brief. (0)</p>	<p>elements are missing. Answer does not fall within the prescribed word count (50 words over the word count). (2)</p>	<p>information is provided and relevant. Answer falls within the prescribed word count. (4)</p>	<p>comprehensive and relevant. Answer falls within the prescribed word count. (6)</p>
<p><b>Question 1:</b> <b>Evidence of understanding and accurate use of the module's content</b></p> <p><i>The answer demonstrates that the student has engaged with the content.</i></p> <p><i>The answer demonstrates that the student has an informed grasp of the advantages and disadvantages of neural networks in comparison to other machine learning models.</i></p>	<p>The answer demonstrates that the student did not engage with the content. OR The answer fails to demonstrate a basic understanding of the content. (0)</p>	<p>The answer demonstrates that the student engaged with the content. The answer demonstrates an inadequate understanding of the content. (2)</p>	<p>The answer demonstrates that the student engaged with the content and understands most of it. (4)</p>	<p>The answer demonstrates that the student engaged with the content and has an excellent understanding of the module's content. (6)</p>
<p><b>Question 1:</b> <b>Coherence and clarity</b></p> <p><i>The answer is clearly structured and written in a way that is comprehensible.</i></p>	<p>Answer is incoherent or lacks clarity. Answer is not logically structured, or is incomprehensible. (0)</p>	<p>Answer shows limited coherence and clarity. The writing is comprehensible but lacks logical structure. (2)</p>	<p>Answer is written clearly and coherently. The writing is logically structured, but there remains some room for improvement. (4)</p>	<p>Answer is extremely well structured and written with exceptional clarity and coherence. (6)</p>
<p><b>Question 2.1:</b> <b>Adherence to the brief</b></p> <p><i>The answer substantiates the student's answers to the different questions based on</i></p>	<p>No submission. OR Answer fails to adhere to any of the elements contained in the brief. (0)</p>	<p>Answer adheres to some elements contained in the brief, but some key elements are missing. Answer does not fall within the prescribed word</p>	<p>Answer adheres to most, but not all, elements of the brief. Almost all information is provided and relevant.</p>	<p>Answer adheres to all elements of the brief. All information provided is comprehensive and relevant. Answer falls within the prescribed word count. (6)</p>

<i>the output of the IDE notebook.</i>		count (50 words over the word count). (2)	Answer falls within the prescribed word count. (4)	
<b>Question 2.2:</b> <b>Adherence to the brief</b> <i>The student substantiates which model's predictions they would use to make a decision.</i>	No submission. OR Answer fails to adhere to any of the elements contained in the brief. (0)	Answer adheres to some elements contained in the brief, but some key elements are missing. Answer does not fall within the prescribed word count (50 words over the word count). (2)	Answer adheres to most, but not all, elements of the brief. Almost all information is provided and relevant. Answer falls within the prescribed word count. (4)	Answer adheres to all the elements of the brief. All information provided is comprehensive and relevant. Answer falls within the prescribed word count. (6)
<b>Question 2:</b> <b>Evidence of understanding and accurate use of the module's content</b> <i>The answer demonstrates that the student has engaged with the content.</i> <i>The answer demonstrates that the student has an informed grasp of the accuracy of both the neural network and the regression model, and how the RMSE value dictates the best choice of model.</i>	Answer demonstrates that the student did not engage with the content. OR Answer fails to demonstrate a basic understanding of the content. (0)	Answer demonstrates that the student engaged with the content. Answer demonstrates an inadequate understanding of the content. (2)	Answer demonstrates that the student engaged with the content and understands most of it. (4)	Answer demonstrates that the student engaged with the content and has an excellent understanding of the module's content. (6)
<b>Question 2:</b> <b>Coherence and clarity</b> <i>The answer is clearly structured and written in a way that is comprehensible.</i>	Answer is incoherent or lacks clarity. Answer is not logically structured, or is incomprehensible. (0)	Answer shows limited coherence and clarity. The writing is comprehensible but lacks logical structure. (2)	Answer is written clearly and coherently. The writing is logically structured, but there remains some room for improvement. (4)	Answer is extremely well structured and written with exceptional clarity and coherence. (6)

**Total: 48 marks**



---

## Feedback:

Start writing here: