Fachhochschule Südwestfalen
University of Applied Sciences

Machine Learning

Project Task 1 Report
within the context of the Machine Learning module
Summer Semester 2024

Professors: Prof. Dr. Andreas De Vries
Prof. Dr. Stefan Böcker

Submitted by: Seyedehsadaf Edalat
Matriculation number: 30372928

Aziz Fahmi
Matriculation number: 30461097

Deadline: 2024-05-27

Wir geben Impulse

Table of Contents

List of Figures

List of Abbreviations

BIC Bayesian Information Criterion

MAU Monthly Active Users

MU Marginal Utility

NA Not Applicable

SSR Sum of Squared Residuals

1 Description of the Problem

The project explores the complex relationship that exists between Meta Platforms' (previously Facebook) total assets from 2008 to 2023 and its monthly active users (MAU). Finding the best model function to predict total assets based on MAU is the main goal. We examine two different models: a linear regression function and the Metcalfe utility function. According to the Metcalfe utility function, which is frequently used in network theory, utility is determined by the product of the user base and one less than the user base, divided by two. This implies that the utility of a network increases quadratically with the number of users, capturing the idea of network effects where the value of the network grows exponentially as more users join. In this case, 'x' denotes the total number of nodes in the network, or the MAU count, and 'a' denotes a scaling coefficient representing the intrinsic value and dynamics of the network. On the other hand, the linear regression function takes the assumption that MAU and total assets have a linear connection, and the slope of the function indicates the pace at which assets change for each new user.

Through this evaluation, we hope to shed light on the valuation mechanisms of Meta Platforms and determine how well each model captures the dynamics of value within the network. In addition, the study aims to clarify how the behaviour of the selected model relates to the idea of declining marginal utility, which is a cornerstone of economics. According to the theory of diminishing marginal utility, the incremental satisfaction gained from consuming more units of an item or service decreases with each new unit. As a result, examining how the chosen model links to this idea might provide insightful information on how Meta Platforms' value changes as its user base grows.

2 Description of the Solution

2.1 Data Import and Cleaning

To begin our analysis, we first imported the data from Meta Platforms' annual reports, which includes the number of monthly active users (MAU) and total assets for the financial years 2008 to 2023. The data was read into R using the `read.csv` function.

FIGURE 1: IMPORTING DATA

```
# Load the CSV data into RStudio
Facebook <- read.csv("C:/Users/sadaf/Downloads/Facebook.csv", sep = "\t", header = TRUE)

# Display the data
print(Facebook)
```

An initial inspection revealed some empty columns and potential irrelevant values that needed to be addressed. Next, we cleaned the data by removing extra columns containing NA values. This step was crucial to ensure the accuracy and reliability of our subsequent analysis. We used the `which` function to identify columns with missing values and removed them from the dataset.

FIGURE 2: CLEANING DATA

```
# Remove extra columns with NA values
Facebook_clean <- Facebook[, -which(colSums(is.na(Facebook)) > 0)]
```

2.2 Scatter Plot

After cleaning the data, we created a scatter plot to visualise the relationship between MAU and total assets. This initial plot provided a visual representation of the data points and helped us understand the general trend and correlation between the two variables. The scatter plot was generated using the `plot` function in R, with MAU on the x-axis and total assets on the y-axis.

FIGURE 3: MAU AGAINST TOTAL ASSETS PLOT

```
# Plot MAU against Total Assets
plot(Facebook_clean$MAU, Facebook_clean$Total.Assets,
     xlab = "Monthly Active Users (MAU)",
     ylab = "Total Assets",
     main = "MAU vs Total Assets Scatter Plot",
     col = "blue", pch = 16)
```

2.3 Linear Regression Model

To model the relationship between MAU and total assets, we first fitted a linear regression model. It assumes a linear relationship between the independent variable (MAU) and the dependent variable (total assets). The fitted linear regression model for our data has the form:

Total Assets=$\beta 0+\beta 1\times$MAU [1]

Where:

- $\beta 0$ is the intercept, representing the total assets when MAU is zero.
- $\beta 1$ is the slope, representing the change in total assets for each additional MAU.

We used the `lm` function in R to fit the linear regression model and added the regression line to our scatter plot.

FIGURE 4: FITTING LINEAR REGRESSION

```
# Fit a linear regression model
linear_model <- lm(Total.Assets ~ MAU, data = Facebook_clean)
```

2.4 Metcalfe Utility Function

Next, we explored the Metcalfe utility function, which is often used to model the value of a network. This nonlinear function accounts for the network effects and is given by:

$uM(x)=a\cdot x\cdot(x-1)/2$

Where:

- x is the number of users (MAU)
- a is a scaling coefficient optimised to fit the data

To fit the Metcalfe utility function to our data, we needed to determine the optimal value of the scaling coefficient *a*. We defined a function to calculate the sum of squared residuals (SSR) between the actual total assets and the predicted values from the Metcalfe utility function. The `optim` function in R was used to find the value of *a* that minimises the SSR.

FIGURE 5: DEFINING METCALFE UTILITY FUNCTION

```r
# Define the Metcalfe utility function
metcalfe_utility <- function(x, a) {
  return(a * x * (x - 1) / 2)
}

# Define a function to calculate sum of squared residuals
ssr <- function(a) {
  predicted_assets <- metcalfe_utility(Facebook_clean$MAU, a)
  return(sum((predicted_assets - Facebook_clean$Total.Assets)^2))
}

# Optimize for the parameter 'a'
optimized_a <- optim(par = 1, fn = ssr)$par
```

After finding the optimal *a*, we added the Metcalfe utility function curve to our scatter plot.

FIGURE 6: ADDING METCALFE UTILITY FUNCTION TO PLOT

```r
# Add Metcalfe utility function curve to the plot with optimized 'a'
curve(metcalfe_utility(x, a = optimized_a), add = TRUE, col = "black")
```

3 Results and Discussions

3.1 Preferred Model Function

We were able to obtain two models—one based on the Metcalfe utility function and the other on linear regression—to estimate Meta Platforms' total assets as a function of its MAU after implementing the analysis using RStudio. The following equation was produced by the linear regression model:

Total Assets = -47027.978 + 72.52×MAU

With an R-squared value of 0.8982, the linear regression model's summary statistics demonstrated a good degree of fit and the capacity to explain 89.1% of the variability in total assets by the number of MAUs. The statistical significance of the association between MAUs and total assets is indicated by the model's substantial p-value of 2.49e-08.

FIGURE 7: SUMMARY OF LINEAR REGRESSION FUNCTION

```
Call:
lm(formula = Total.Assets ~ MAU, data = Facebook_clean)

Residuals:
   Min     1Q Median     3Q    Max 
-24142 -19934  -4350   8904  54351 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -47027.978  12885.208   -3.65  0.00263 ** 
MAU             72.529      6.526   11.11 2.49e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 24880 on 14 degrees of freedom
Multiple R-squared:  0.8982,	Adjusted R-squared:  0.8909 
F-statistic: 123.5 on 1 and 14 DF,  p-value: 2.489e-08
```

The Metcalfe utility function yielded the parameter 'a' that best fits the given data after being optimised using the sum of squared residuals.

Following optimization, the quadratic model's summary statistics (which include the Metcalfe utility function) revealed an even higher R-squared value of 0.9853, indicating that the MAUs in this model can account for almost 98.3% of the variability in total assets. The model exhibited a noteworthy p-value of 1.24e-12, signifying a robust correlation between MAUs and total assets.

Figure 8: Summary of Metcalfe Utility Function

```
Call:
lm(formula = Total ~ MAU + I(MAU^2), data = Facebook)

Residuals:
     Min       1Q   Median       3Q      Max 
-16352.4  -3896.6   -565.9   2443.9  24835.4 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) 4688.82152 7789.03861   0.602   0.5575    
MAU          -20.42844   10.91184  -1.872   0.0839 .  
I(MAU^2)       0.02797    0.00319   8.767  8.1e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9821 on 13 degrees of freedom
Multiple R-squared:  0.9853,	Adjusted R-squared:  0.983 
F-statistic: 434.8 on 2 and 13 DF,  p-value: 1.24e-12
```

Another criterion that helped with the conclusion was Bayesian Information Criterion (BIC). It is based on the likelihood function and includes a penalty term for the number of parameters in the model to discourage overfitting. The formula for BIC is: $BIC = k\log(n) - 2\log(L)$

where:

- k is the number of parameters
- n is the number of observations
- L is the likelihood of the model [2]

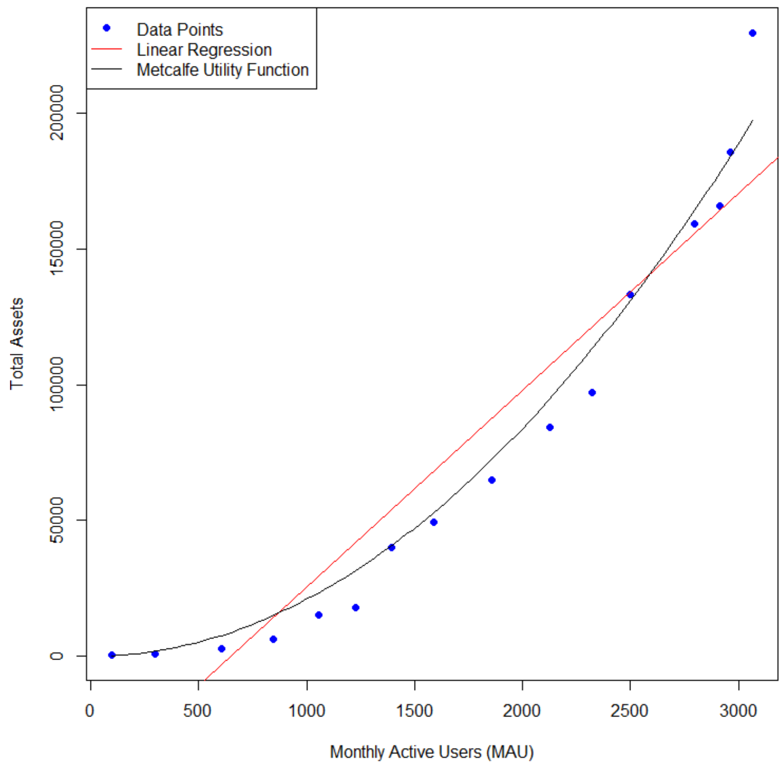
In our analysis, the BIC for the linear regression model was 375.4874. And the BIC for the Metcalfe utility function model was 299.1493. A lower BIC value indicates a better model, taking into account both the goodness of fit and the complexity of the model.

To summarise the comparison, let’s view from different perspectives:

Fit to Data: The Metcalfe utility function provided a better fit to the data, as indicated by the higher adjusted R-squared value and lower residual standard error compared to the linear regression model.

Theoretical Justification: The Metcalfe utility function is grounded in network theory, which posits that the value of a network increases with the square of the number of users due to the increasing number of possible connections.

FIGURE 9: MAU VS TOTAL ASSETS SCATTER PLOT



3.2 Relation to the Law of Diminishing Marginal Utility

The law of diminishing marginal utility states that as the quantity of a good or service consumed increases, the additional satisfaction (utility) gained from consuming an additional unit decreases. [3]

The linear regression function implies a constant marginal increase in total assets for each additional user. This does not align with the law of diminishing marginal utility, as it suggests that each additional user provides the same incremental value to the network regardless of the current number of users.

The Metcalfe utility function suggests that the value of the network grows quadratically with the number of users. Initially it seems to go against diminishing marginal utility because it shows a big increase in value with each new user. While Metcalfe's Law predicts quadratic growth in network utility, real-world limitations and behavioural factors lead to diminishing returns as the network becomes extremely large. Thus, over time, the growth in utility **may** start to align with the law of diminishing marginal utility, although the current data does not show that. (see appendix)

Bibliography

[1] Dr. De Vries, A. (2024): *Lecture notes on Machine Learning*

[2] Brownlee, J. (2020): *Probabilistic Model Selection with AIC, BIC, and MDL.* https://machinelearningmastery.com/probabilistic-model-selection-measures/

[3] Kenton, W. (2024): *The Law of Diminishing Marginal Utility: How It Works, With Examples*. https://www.investopedia.com/terms/l/lawofdiminishingutility.asp

Appendix

For the Metcalfe utility function, the marginal utility for the user X can be calculated as:

MU(x) = uM(x) − uM(x−1). For example, if optimized_a is approximately 0.028, then:

MAU	Total Utility (uM)	Marginal Utility
100	1386	1386
300	12573	11187
608	51589	39016
845	99781	48192
1056	155736	55955
1228	210318	54582
1393	271781	61463
1591	353545	81764
1860	482184	128639
2129	627557	145373
2320	807680	180123
2498	987215	179535
2797	1350357	363142
2912	1548532	198175
2963	1640841	92309
3065	1840701	199860