Machine Learning

Project 01

28.05.2024

Seyedehsadaf Edalat (30372928)
Aziz Fahmi (30461097)

Table of Contents

Introduction

Meta Platforms Overview

- Meta Platforms, Inc., formerly Facebook, is a global leader in social media, connecting billions of users through platforms like Facebook, Instagram, and WhatsApp.

Importance of Network Value

- Estimating Meta's network value helps investors, analysts, and stakeholders understand the company's financial health and growth potential. It's also vital for strategic decisions and assessing economic impact.

Project Goals

- This project aims to determine the best model to estimate Meta's total assets based on Monthly Active Users (MAU), comparing linear regression and the Metcalfe utility function, and exploring their economic implications.

Project Problem

?

Which model better estimates total assets:
the Metcalfe utility function or linear regression ?

?

How does the chosen model relate to the law of diminishing marginal utility?

Data Description

Data source and period

Meta Platforms' annual reports
Financial years: 2008 to 2023

Key variables

Monthly Active Users (MAU): Number of active users each month
Total Assets: The value of Meta's assets

Data cleaning

Removed columns with NA values

Methodology

Linear Regression Model

Fits a linear relationship between MAU and Total Assets

Total Assets: B0+B1*MAU

Metcalfe Utility Function

Based on network theory Um(x)=a*x*(x-1)/2

Optimized parameter a to fit the data

Assumptions

Total Assets are influenced by the number of active users. Data points are independent and representative of the overall trend.

Model Evaluation

Compared models using adjusted R-squared, residual standard error, and Bayesian Information Criterion (BIC)

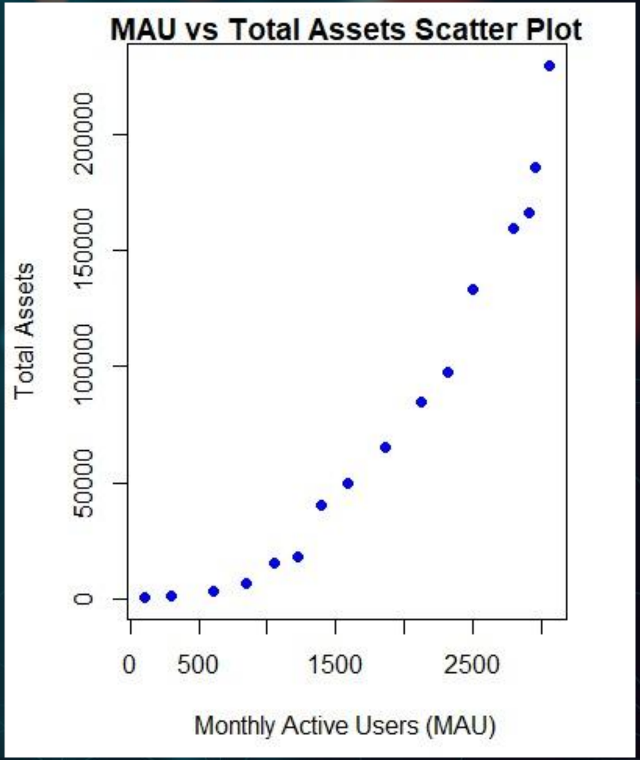
Code Implementation

Load Data and Clean:

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

# Display the data
print(Facebook)

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

Scatter Plot:

```r
# 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)
```



Code Implementation

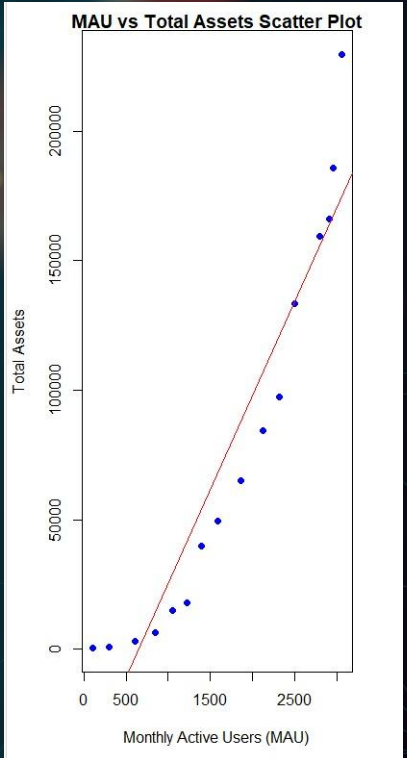
Linear Regression Model:

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

# Add linear regression line to the plot
abline(linear_model, col = "red")

```

These lines fit a linear regression model to the cleaned Facebook data and add the resulting regression line to the scatter plot.

Method: Using lm function to fit a linear regression model to the relationship between the total assets and the monthly active users .

Data: Utilizes the Facebook_clean dataset containing non-missing values for both Total.Assets and MAU.



Code Implementation

- **Metcalfe Utility Function:**
  - Equation: $uM(x)=a\cdot x\cdot(x-1)/2$
  - *x*: Number of Monthly Active Users (MAU)
  - *a*: Parameter optimized to best fit the data

```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))
}
```

- Optimizing Parameter *a*:
  - Objective: Minimize Sum of Squared Residuals (SSR)
  - SSR measures the difference between predicted and actual total assets
  - Initial assumption: a=1
  - Iterative process to find optimal a
  - Result: Optimized *a* value is approximately 0.0419216

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

- Final Metcalfe Utility Function:
  - $uM(x)=0.0419216\cdot x\cdot(x-1)/2$

Code Implementation

Bayesian Information Criterion:

```
# Calculate BIC for linear regression model
BIC_linear <- BIC(linear_model)

# Calculate BIC for Metcalfe utility function model
residuals <- residuals(metcalfe_model)
n <- length(residuals)
k <- length(coef(metcalfe_model))  # Number of parameters including intercept
BIC_metcalfe <- n * log(sum(residuals^2) / n) + k * log(n)

```

Residuals: Obtains the residuals from the Metcalfe utility function model.

n <- length(residuals) calculates the number of observations in the dataset.

K <- length(coef(metcalfe_model)) determines the number of parameters in the model, including the intercept.

Code Implementation

Bayesian Information Criterion:

```
BIC_metcalfe <- n * log(sum(residuals^2) / n) + k * log(n)
```

Sum of Squared Residuals (SSR) = sum(residuals^2)

Mean Squared Residuals: Sum(residuals^2) / n gives the average of the squared residuals.

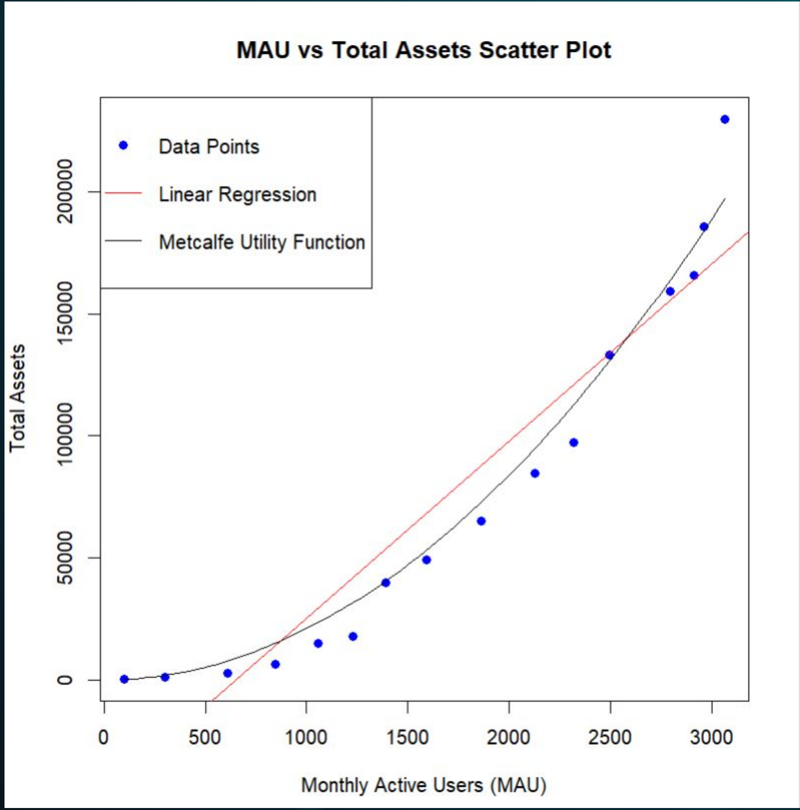
Log-Likelihood Part (n * log(sum(residuals^2) / n)): This part represents the goodness of fit. It penalizes the model based on how well it fits the data. Lower values indicate a better fit.

Penalty for Model Complexity (k * log(n)): This part penalizes the model based on the number of parameters (complexity). Adding more parameters usually improves the fit, but this term penalizes unnecessary complexity to prevent overfitting.

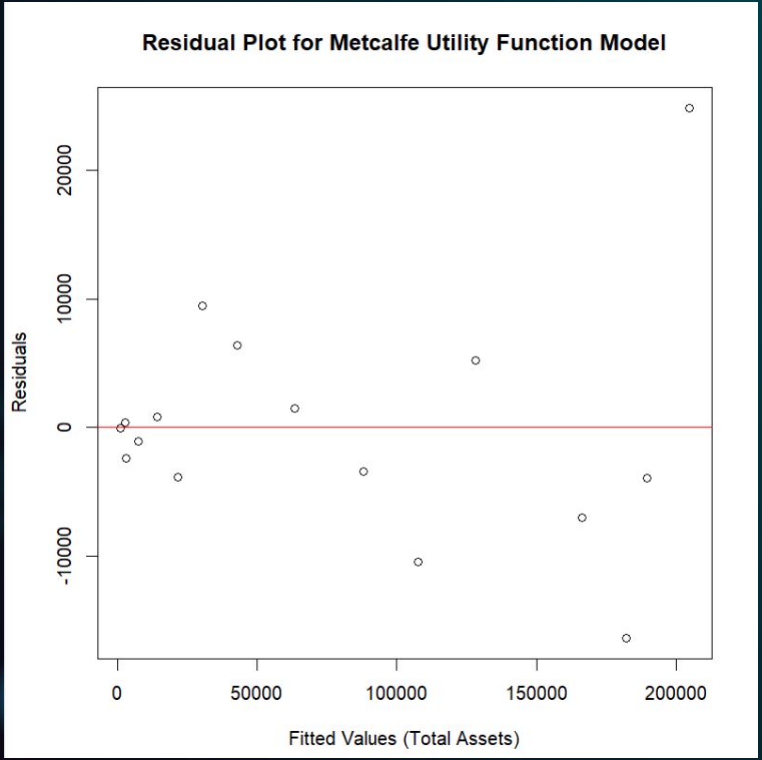
Results

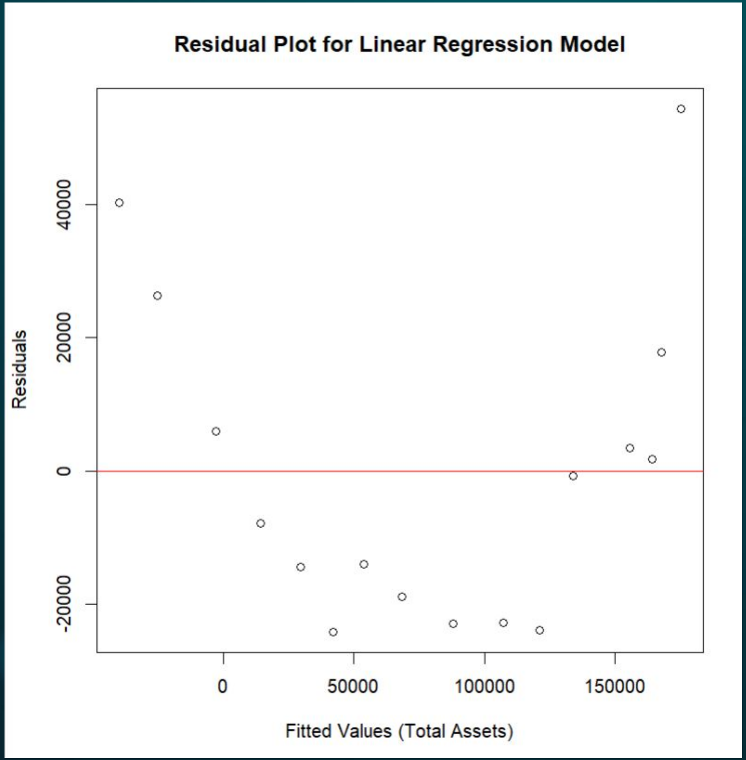
Linear Regression	Metcalfe
• Adjusted R-squared: 0.8909 • Multiple R-squared: 0.8982 • Residual Standard Error: 24880 • BIC: 375.4874 • P_value: 2.489e-08	• Adjusted R-squared: 0.983 • Multiple R-squared: 0.9853 • Residual Standard Error: 9821 • BIC: 299.1493 • P_value: 1.24e-12
Equation: **Total Assets = -47027.978 + 72.529 * MAU**	**Equation:** **Total Assets = 0.0419216×MAU×(MAU−1)/2**

Visualisation



Visualisation





Law of Diminishing Marginal Utility

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

This law states that as you consume more of something, each additional unit gives you less added satisfaction.

Linear Regression Function:

Assumes a straight-line relationship between users (MAU) and total assets. Every new user adds the same amount of value. This doesn't fit with diminishing marginal utility because it suggests each new user is always equally valuable.

Metcalfe Utility Function:

Suggests the value of the network increases quickly as more users join (quadratically). 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 behavioral 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)

THANK YOU!