سلطنة عُمان
وزارة النقل والاتصالات وتقنية المعلومات
Sultanate of Oman
Ministry of Transport, Communications and Information Technology

# Insights into Linear and Logistic Regression

**Trainee:** Iman Abdullah Al-Battashi

**Instructor:** Kulsoom Shoukat

# Table of Contents



# Table of figures:

# Introduction:

Linear Regression and Logistic Regression are Supervised Machine Learning models that employ labeled datasets to predict outcomes. However, there is an important difference between their applications: Linear Regression is utilized for regression issues in which the aim is to anticipate a continuous numeric variable, such as home prices or temperature. Logistic Regression is typically used to solve categorical classification issues, such as detecting whether an email is spam (yes/no) or if a buyer would purchase a product (1/0).

## Linear Regression:

### What is Linear Regression:

Linear regression is a supervised machine learning algorithm that predicts continuous numerical values by assuming a linear connection between one or more input characteristics (independent variables) and an output (dependent variable). The algorithm creates a straight line from the data to best reflect how changes in the input impact the result.
It is commonly used in predictive analytics, which aims to predict outcomes like as sales, pricing, or trends based on present behaviors. Linear regression asserts that the dependent variable varies in proportion to the independent variable(s), making it useful for predicting and trend analysis in fields such as finance, economics, and science. (Kanade, 2025)

### Key advantages of Linear Regression

- **Simple and efficient:** Easy to implement with low computational demands, making it quick to train and maintain.
- **Highly interpretable:** Clearly shows how each input variable affects the prediction, offering transparent insights.
- **Scalable:** Handles large datasets well without requiring heavy resources.
- **Real-time friendly:** Can be updated quickly with new data, ideal for online and real-time applications.

### Types of Linear Regression:

Linear regression has several types you can use depending on the data and problem. Some common types include:

- **Simple Linear Regression**
  Models the relationship between one independent and one dependent variable.
  ***Example****:* Predicting pollution levels based on temperature.
- **Multiple Linear Regression**
  Uses two or more independent variables to predict a continuous outcome.
  ***Example****:* Estimating blood pressure from height, weight, and exercise habits.
- **Ordinal Regression**
  Handles dependent variables with ordered categories using predictor variables.
  ***Example****:* Survey answers rated as disagree, neutral, or agree.

**Mathematical Foundation of Linear Regression**

Linear Regression models the relationship between input x and output y using:

$$y(x) = p_0 + p_1 x$$

Where:

• $p_0$: intercept (bias)
• $p_1$: coefficient (slope)
• x: input feature
• y: predicted output

For multiple features:

$$y(x) = p_0 + p_1 x_1 + p_2 x_2 + \dots + p_n x_n$$

**Cost Function – MSE**

The model minimizes Mean Squared Error (MSE):

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (y_i - (p_1 x_i + p_0))^2$$

Where $y_i$ is the actual value and $x_i$ is the input.

**Example Plots:**

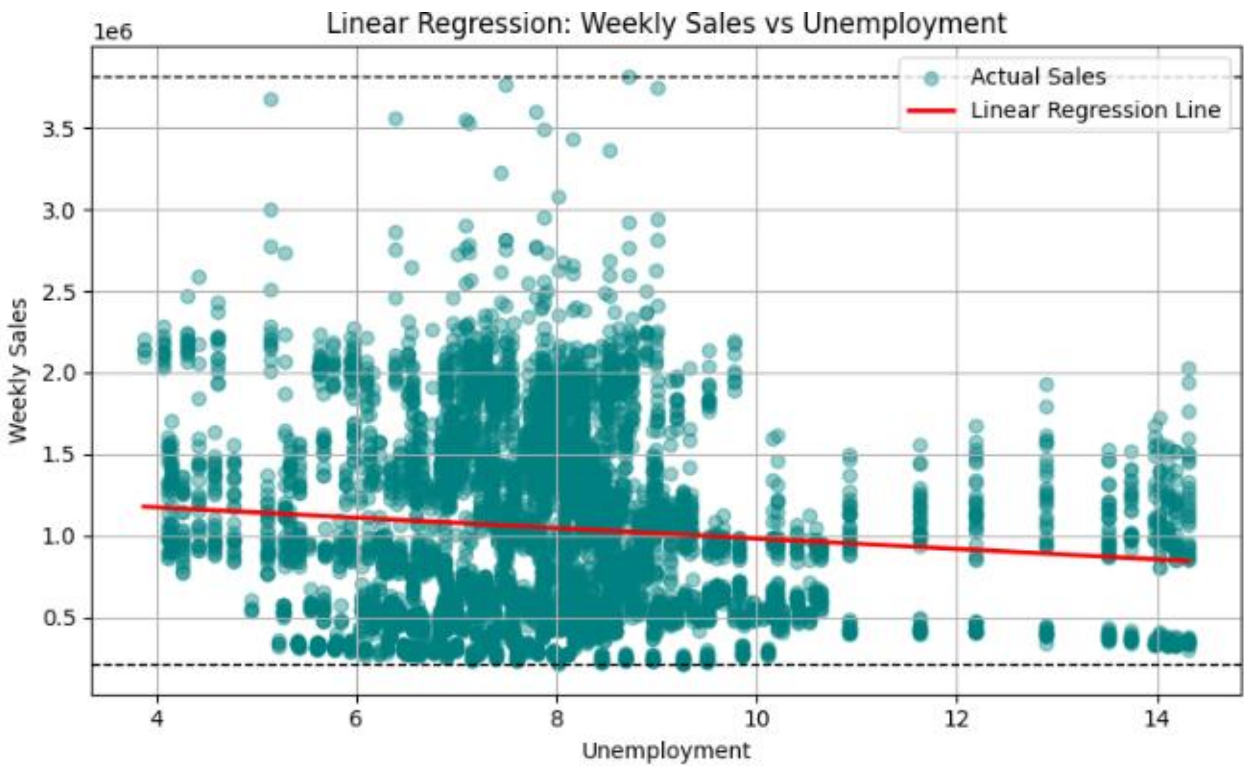


*Figure 1: Linear Regression Plot*

The plot illustrates a Linear Regression model that examines the association between unemployment rates and weekly sales. Each teal dot indicates a weekly sales record that corresponds to the given unemployment rate. The red line represents the linear regression model's projection of weekly sales based on unemployment rates. The negative slope of the line implies a minor inverse relationship: as unemployment rises, weekly sales fall. However, the broad range of data points around the line suggests that unemployment alone does not substantially predict sales, and that other factors may have a larger influence.

## Logistic Regression:

**What is logistic regression:**

Logistic regression is a supervised machine learning algorithm that usually deals with binary classification problems. It operates by evaluating the likelihood that a given input falls into one of two categories—yes or no, 0 or 1, or true or false. The final prediction is formed by assigning this probability to a given class based on a threshold, usually 0.5.

This method evaluates the link between one or more independent factors and a categorical result, allowing it to divide data into various categories. It is commonly used in predictive analytics for determining the chance that a given incident falls into a specified category. (Kanade, 2025a).

**Key advantages of logistic Regression:**

- **Easy to implement:** Lightweight and fast to train, requiring minimal computational power.
- **Effective for binary classification:** Works well when classes are linearly separable, providing reliable predictions.
- **Insightful coefficients:** Quantifies the influence and direction of predictors on the outcome, aiding interpretability.

**Types of Logistic Regression:**

Logistic regression has different forms depending on the outcome variable. The main types are:

1. **Binary Logistic Regression**
   Predicts an outcome with **two possible classes** (e.g., yes/no).
   **Example:** Predicting if a bank customer will default on a loan (Yes or No).

2. **Multinomial Logistic Regression**
   Used when the outcome has **more than two categories with no order**.
   **Example:** Predicting a student's choice of college major: Science, Arts, or Commerce.

3. **Ordinal Logistic Regression**
   For outcomes with **ordered categories**.
   **Example:** Survey responses rated as Disagree, Neutral, or Agree.

**Mathematical Foundation of Logistic Regression**

Used for classification, Logistic Regression uses the sigmoid function:

$$f(x) = \frac{1}{1+e^{-x}}$$

The logistic model:

Where:

- x: input
- y: predicted probability
- $b_0$: bias
- $b_1$: weight

$$y = \frac{e^{b_0 + b_1 x}}{1 + e^{b_0 + b_1 x}}$$

**Example Plot:**

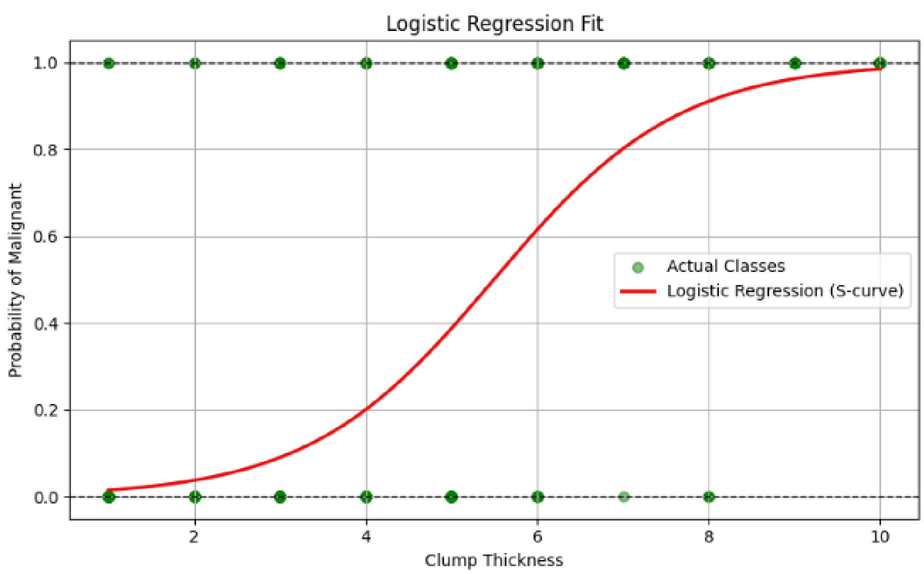


*Figure 2: Logistic Regression Plot*

The graphic above shows the outcome of using Logistic Regression to estimate the probability of a tumor being malignant using the attribute Clump Thickness. The green scatter dots indicate the real class designations (0 benign, 1 malignant). The red S-shaped curve is the logistic regression model's prediction, demonstrating how the chance of malignancy grows with clump thickness. This S-curve highlights the nonlinear character of logistic regression, which makes it ideal for binary classification tasks such as cancer diagnosis. The dashed horizontal lines at 0 and 1 represent the probability value boundaries.

# Conclusion:

This research used real-world datasets to compare and evaluate linear and logistic regression. Linear regression accurately predicted Walmart's weekly sales by modeling relationships with continuous variables, indicating its utility in numerical prediction problems. Logistic regression was used on breast cancer data to demonstrate its capabilities in binary classification by discriminating between benign and malignant instances.

The study used theoretical principles, mathematical foundation, and visuals to clearly demonstrate the performance and applicability of each model. Overall, this study gives a thorough grasp of regression techniques, their benefits, and practical applications in prediction and classification issues.

# References


1. Kanade, V. (2025a, March 10). *Everything you need to know about logistic regression*. Spiceworks Inc. https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-logistic-regression/

2. Kanade, V. (2025b, March 10). *What is linear regression?* Spiceworks Inc. https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-linear-regression/

3. Shut, M. (2023, January 8). *Breast cancer (cleaned)*. Kaggle. https://www.kaggle.com/datasets/marshuu/breast-cancer
4. Mikhail. (2024, June 6). *Walmart sales*. Kaggle. https://www.kaggle.com/datasets/mikhail1681/walmart-sales?resource=download