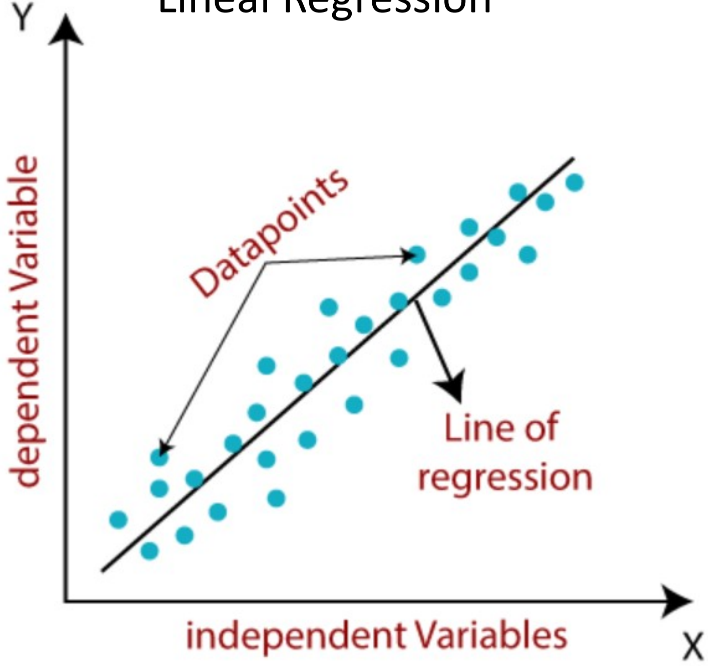
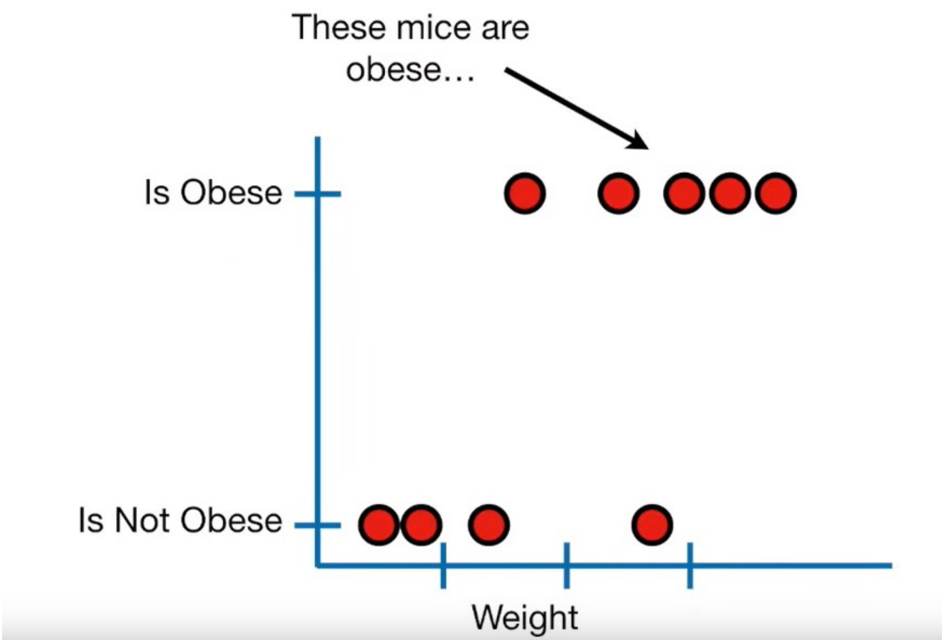
Logistic regression

Abdellah El Fallahi

Lineal Regression



Logistic Regression



Agenda

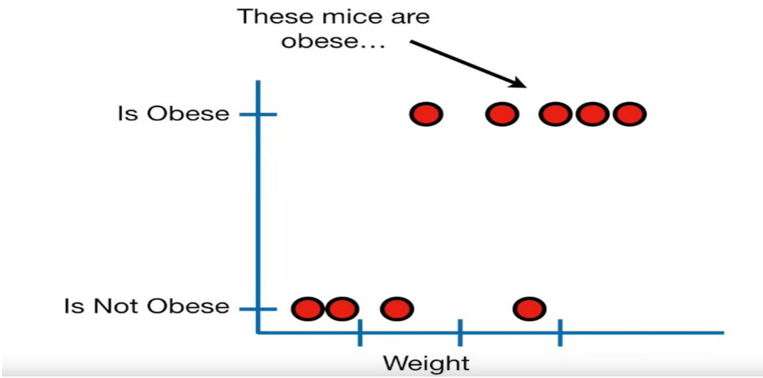
1. Introduction
2. Logistic Regression Overview
3. Mathematical Foundation
4. Sigmoid Function
5. Cost Function
6. Gradient Descent
7. Assumptions
8. Model Building Steps
9. Sample Dataset
10. Data Preprocessing
11. Model Training
12. Model Evaluation
13. Interpretation
14. Applications
15. Conclusion

Introduction

- Logistic regression is a statistical method used for binary classification.

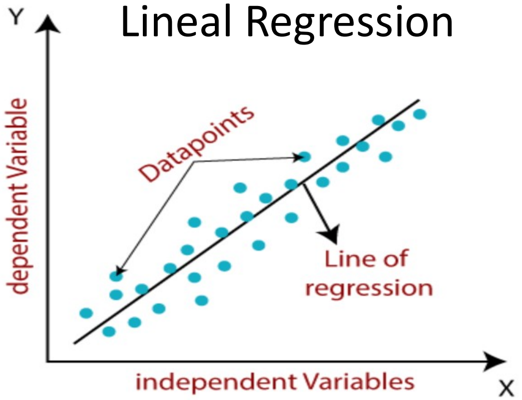
- It predicts the probability that an instance belongs to a particular class.

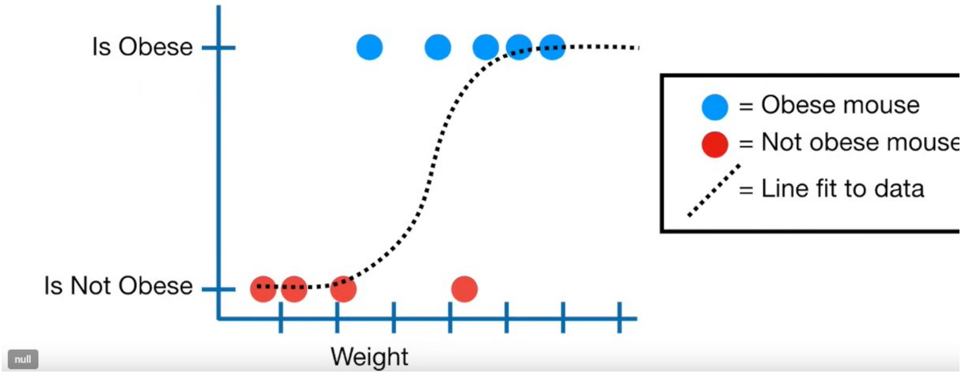
Logistic Regression



Logistic Regression Overview

- Unlike linear regression, logistic regression outputs probabilities and maps them to classes using a threshold.





Lineal Regression Vs Logistic regression

Loss function LR

- Minimiser la fonction:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2$$

Loss Function Log Regression

$$P(y = 1|X) = \text{sigmoid}(z) = \frac{1}{1 + e^{-z}}$$

where

$$z = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \ldots + \hat{\beta}_k x_k$$

On chercher a minimiser la fonction Log-Loss

A suitable loss function in logistic regression is called the Log-Loss, or binary cross-entropy. This function is:

$$\text{Log-Loss} = \sum_{i=0}^{n} -(y_i * \log(p_i) + (1 - y_i) * \log(1 - p_i))$$

Min Log-Loss est équivalent a maximiser Log-Likelihood

$$\text{Log-Likelihood} = \sum_{i=0}^{n} (y_i * \log(p_i) + (1 - y_i) * \log(1 - p_i))$$

Termenology

$$Odds = \frac{Probability\ Event\ Occurs\ (p)}{Probability\ Event\ Does\ Not\ Occur\ (1-p)}$$

In logistic regression, we model the **probability** that the output $Y = 1$ given the input X:

$$P(Y = 1 \mid X) = \hat{p}(X)$$

This probability is modeled using the **logistic (sigmoid) function**:

$$\hat{p}(X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n)}}$$

Or more compactly:

$$\hat{p}(X) = \frac{1}{1 + e^{-z}} \quad \text{where } z = \beta_0 + \sum_{i=1}^{n} \beta_i x_i$$

Odds

The **odds** of an event are defined as the ratio of the probability the event happens to the probability it doesn't:

$$\text{Odds} = \frac{P(Y = 1 \mid X)}{1 - P(Y = 1 \mid X)} = \frac{\hat{p}(X)}{1 - \hat{p}(X)}$$

Logistic regression models the **log of the odds** (called the *logit*) as a linear function of the input variables:

$$\log\left(\frac{\hat{p}(X)}{1 - \hat{p}(X)}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n$$

This is the core of logistic regression: we **predict log-odds** linearly, then use the sigmoid to get the actual probability.

Concept	Formula
Probability	$\hat{p}(X) = \frac{1}{1+e^{-z}}$
Odds	$\frac{\hat{p}(X)}{1-\hat{p}(X)}$
Log-Odds	$\log\left(\frac{\hat{p}(X)}{1-\hat{p}(X)}\right) = \beta_0 + \sum \beta_i x_i$

Interpretation of Log Regression Model

A Binary Feature | A Continuous Feature | Multivariate Regression

Interpreting A Logistic Regression Model With One Binary Feature

Model Form: $P(y=1|x) = \frac{1}{1+e^{-(\hat{\beta}_0+\hat{\beta}_1 * x)}}$ where $x \in [\![0,1]\!]$

Example: $P(Day = Sunny|Foggy) = \frac{1}{1+e^{-(\hat{\beta}_0+\hat{\beta}_1 * Foggy)}}$

Interpretation: This model summarizes the difference in the probability of a sunny day for days that are foggy and days that are not.

The weight $\hat{\beta}_1$ is the change in the log-odds ratio for a foggy day relative to a non-foggy day. For a coefficient of -0.7, the exponentiated value is $e^{-0.7} = 0.50$, which indicates that on average, the odds that the day will be sunny multiplies by 0.50 (i.e. is about half) if it is foggy compared to if it is not.

The intercept $\hat{\beta}_0$ is the odds of a sunny day if it is not foggy.

Interpretation of Log Regression Model

A Binary Feature | A Continuous Feature | Multivariate Regression

Interpreting A Logistic Regression Model With One Continuous Feature

Model Form: $P(y=1|x) = \frac{1}{1+e^{-(\hat{\beta}_0+\hat{\beta}_1 * x)}}$ where $x \in [\![\mathbb{R}]\!]$

Example: $P(Day = Sunny|Temperature) = \frac{1}{1+e^{-(\hat{\beta}_0+\hat{\beta}_1 * Temperature)}}$

Interpretation: This model describes the probability of a sunny day across temperatures (measured in degrees Fahrenheit).

The weight $\hat{\beta}_1$ is the change in the log-odds ratio for y per unit change in X. When exponentiating the weight, this becomes the change in the odds ratio for y per unit change in X. In other words, the odds are multiplied by $e^{\hat{\beta}_1}$. Therefore, for a coefficient of 0.7, the exponentiated value is $e^{0.7} = 2.01$, which indicates that on average, for an increase of 1 degree in temperature, the odds that the day will be sunny multiplies by 2.01 (i.e. approximately doubles).

The intercept $\hat{\beta}_0$ is the odds of a sunny day at 0 degrees Fahrenheit. We can use this to calculate the probability of a sunny day by using the following calculation: $p = \frac{e^{\hat{\beta}_i}}{1+e^{\hat{\beta}_i}}$.

If the intercept coefficient is 2, this means that there is a $p = \frac{e^2}{1+e^2} = 0.88$ probability that it will be a sunny day at 0 degrees Fahrenheit.

Interpretation of Log Regression Model

A Binary Feature | A Continuous Feature | Multivariate Regression

Interpreting A Multivariate Logistic Regression Model

Model Form: $P(y=1|x) = \frac{1}{1+e^{-(\hat{\beta}_0+\hat{\beta}_1*x_1+\hat{\beta}_2*x_2+...+\hat{\beta}_n*x_n)}}$

Example: $P(Day = Sunny|Temperature, Foggy) = \frac{1}{1+e^{-(\hat{\beta}_0+\hat{\beta}_1*Temperature+\hat{\beta}_2*Foggy)}}$

Interpretation: Typically, a logistic regression model will contain more than one feature. We call this a *multivariate logistic regression model*. In this example, we model the probability of a sunny day as a function of temperature and whether or not it is foggy.

The intercept represents the predicted probability of a sunny day for Days with all $x_i = 0$: it represents the probability for days without fog and with a temperature of zero degrees Fahrenheit.

The weight $\hat{\beta}_1$ is the change in the log-odds ratio for a sunny day per unit change in temperature, and the weight $\hat{\beta}_2$ is the change in the log-odds ratio for a foggy day relative to a non-foggy day.

Confusion matrix

- A **confusion matrix** is a table used to **evaluate the performance** of a classification model, especially in binary classification.

	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

Metrics obtained from the CM

🧮 From the Confusion Matrix, You Can Get:

- **Accuracy** = (TP + TN) / (TP + FP + TN + FN)
- **Precision** = TP / (TP + FP) → how many predicted positives were actually positive?
- **Recall (Sensitivity, TPR)** = TP / (TP + FN)
- **F1 Score** = 2 × (Precision × Recall) / (Precision + Recall)

Receiver Operating Characteristic: ROC

- It’s a **graphical plot** that shows the performance of a binary classification model (like logistic regression) **at all possible classification thresholds**.

What Does the ROC Curve Show?

- **True Positive Rate (TPR)** = Sensitivity = Recall = TP / (TP + FN)
  → How many actual positives you correctly identified
- **False Positive Rate (FPR)** = FP / (FP + TN)
  → How many actual negatives were incorrectly labeled as positive
- **So:**
- Y-axis: TPR (good)
- X-axis: FPR (bad)

Interpreting the Curve ROC

- A good model **hugs the top-left corner** (high TPR, low FPR)
- A random model is a diagonal line from (0,0) to (1,1)
- A perfect model goes straight up the y-axis to (0,1), then straight across to (1,1)

Area Under the Curve (AUC)

- **AUC** = Area Under the Curve
  It gives you a single number:
- **1.0** → perfect classifier
- **0.5** → random guessing
- **< 0.5** → worse than guessing (something is off)

What do We Mean by **Random Guessing** in Classification?

- When we say a model is "random guessing," it means:
- The model is **not learning anything useful** from the data and is just **randomly picking labels** (like flipping a coin) when is has to classify something.

What do We Mean by **Random Guessing** in Classification?

- **On the ROC Curve:**
- A **random guesser** follows the **diagonal line** from (0,0) to (1,1), where:
  - TPR = FPR at all points
  - That means: for every true positive it finds, it also falsely classifies a negative as positive just as often.
- This gives an **AUC = 0.5**, which is equivalent to **no predictive power**.

Conclusion

- Logistic regression is a powerful tool for binary classification tasks.
- Simple yet effective with interpretable results.