

Logistic Regression in Machine Learning

Logistic regression is a supervised machine learning algorithm mainly used for classification tasks where the goal is to predict the probability that an instance of belonging to a given class or not. It is a kind of statistical algorithm, which analyze the relationship between a set of independent variables and the dependent binary variables. It is a powerful tool for decision-making. For example email spam or not.

Logistic Regression

Logistic regression is a supervised machine learning algorithm mainly used for classification tasks where the goal is to predict the probability that an instance of belonging to a given class. It is used for classification algorithms its name is logistic regression. it's referred to as regression because it takes the output of the linear regression function as input and uses a sigmoid function to estimate the probability for the given class. The difference between linear regression and logistic regression is that linear regression output is the continuous value that can be anything while logistic regression predicts the probability that an instance belongs to a given class or not.

Terminologies involved in Logistic Regression:

Here are some common terms involved in logistic regression:

- Independent variables: The input characteristics or predictor factors applied to the dependent variable's predictions.
- Dependent variable: The target variable in a logistic regression model, which we are trying to predict.
- Logistic function: The formula used to represent how the independent and dependent variables relate to one another. The logistic function transforms the input variables into a probability value between 0 and 1, which represents the likelihood of the dependent variable being 1 or 0.
- Odds: It is the ratio of something occurring to something not occurring. it is different from probability as the probability is the ratio of something occurring to everything that could possibly occur.

- Log-odds: The log-odds, also known as the logit function, is the natural logarithm of the odds. In logistic regression, the log odds of the dependent variable are modeled as a linear combination of the independent variables and the intercept.
- Coefficient: The logistic regression model's estimated parameters, show how the independent and dependent variables relate to one another.
- Intercept: A constant term in the logistic regression model, which represents the log odds when all independent variables are equal to zero.
- Maximum likelihood estimation: The method used to estimate the coefficients of the logistic regression model, which maximizes the likelihood of observing the data given the model.

How does Logistic Regression work?

The logistic regression model transforms the linear regression function continuous value output into categorical value output using a sigmoid function, which maps any real-valued set of independent variables input into a value between 0 and 1. This function is known as the logistic function.

Let the independent input features be

$$X = \begin{bmatrix} x_{11} & \dots & x_{1m} \\ x_{21} & \dots & x_{2m} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{nm} \end{bmatrix}$$

and the dependent variable is Y having only binary value i.e. 0 or 1.

$$Y = \begin{cases} 0 & \text{if } \textit{Class 1} \\ 1 & \text{if } \textit{Class 2} \end{cases}$$

then apply the multi-linear function to the input variables X

$$z = \left(\sum_{i=1}^n w_i x_i \right) + b$$

Here x_i is the i th observation of X , $w_i = [w_1, w_2, w_3, \dots, w_m]$ is the weights or Coefficient, and b is the bias term also known as intercept. simply this can be represented as the dot product of weight and bias.

$$z = w \cdot X + b$$

Sigmoid Function

Now we use the **sigmoid function** where the input will be z and we find the probability between 0 and 1. i.e predicted y .

$$\sigma(z) = \frac{1}{1+e^{-z}}$$

Sigmoid function

As shown above, the figure sigmoid function converts the continuous variable data into the **probability** i.e. between 0 and 1.

- $\sigma(z)$ tends towards 1 as $z \rightarrow \infty$
- $\sigma(z)$ tends towards 0 as $z \rightarrow -\infty$
- $\sigma(z)$ is always bounded between 0 and 1

where the probability of being a class can be measured as:

$$P(y = 1) = \sigma(z)$$

$$P(y = 0) = 1 - \sigma(z)$$

Logistic Regression Equation

The odd is the ratio of something occurring to something not occurring. it is different from probability as the probability is the ratio of something occurring to everything that could possibly occur. so odd will be

$$\frac{p(x)}{1-p(x)} = e^z$$

Applying natural log on odd. then log odd will be

$$\log \left[\frac{p(x)}{1-p(x)} \right] = z$$

$$\log \left[\frac{p(x)}{1-p(x)} \right] = w \cdot X + b$$

then the final logistic regression equation will be:

$$p(X; b, w) = \frac{e^{w \cdot X + b}}{1 + e^{w \cdot X + b}} = \frac{1}{1 + e^{-w \cdot X + b}}$$

Likelihood function for Logistic Regression

The predicted probabilities will $p(X; b, w) = p(x)$ for $y=1$ and for $y = 0$
 predicted probabilities will $1-p(X; b, w) = 1-p(x)$

$$L(b, w) = \prod_{i=1}^n p(x_i)^{y_i} (1 - p(x_i))^{1-y_i}$$

Taking natural logs on both sides

$$\begin{aligned} l(b, w) = \log(L(b, w)) &= \sum_{i=1}^n y_i \log p(x_i) + (1 - y_i) \log(1 - p(x_i)) \\ &= \sum_{i=1}^n y_i \log p(x_i) + \log(1 - p(x_i)) - y_i \log(1 - p(x_i)) \\ &= \sum_{i=1}^n \log(1 - p(x_i)) + \sum_{i=1}^n y_i \log \frac{p(x_i)}{1 - p(x_i)} \\ &= \sum_{i=1}^n -\log 1 - e^{-(w \cdot x_i + b)} + \sum_{i=1}^n y_i (w \cdot x_i + b) \\ &= \sum_{i=1}^n -\log 1 + e^{w \cdot x_i + b} + \sum_{i=1}^n y_i (w \cdot x_i + b) \end{aligned}$$

Gradient of the log-likelihood function

To find the maximum likelihood estimates, we differentiate w.r.t w ,

$$\begin{aligned}\frac{\partial J(l(b, w))}{\partial w_j} &= - \sum_{i=1}^n \frac{1}{1 + e^{w \cdot x_i + b}} e^{w \cdot x_i + b} x_{ij} + \sum_{i=1}^n y_i x_{ij} \\ &= - \sum_{i=1}^n p(x_i; b, w) x_{ij} + \sum_{i=1}^n y_i x_{ij} \\ &= \sum_{i=1}^n (y_i - p(x_i; b, w)) x_{ij}\end{aligned}$$

Assumptions for Logistic Regression

The assumptions for Logistic regression are as follows:

- Independent observations: Each observation is independent of the other. meaning there is no correlation between any input variables.
- Binary dependent variables: It takes the assumption that the dependent variable must be binary or dichotomous, meaning it can take only two values. For more than two categories softmax functions are used.
- Linearity relationship between independent variables and log odds: The relationship between the independent variables and the log odds of the dependent variable should be linear.
- No outliers: There should be no outliers in the dataset.
- Large sample size: The sample size is sufficiently large

Types of Logistic Regression

Based on the number of categories, Logistic regression can be classified as:

Binomial Logistic regression:

target variable can have only 2 possible types: “0” or “1” which may represent “win” vs “loss”, “pass” vs “fail”, “dead” vs “alive”, etc., in this case, sigmoid functions are used, which is already discussed above.

Multinomial Logistic Regression

target variable can have 3 or more possible types which are not ordered(i.e. types have no quantitative significance) like “disease A” vs “disease B” vs “disease C”.

In this case, the softmax function is used in place of the sigmoid function. Softmax function for K classes will be:

$$\text{softmax}(z_i) = \frac{e^{z_i}}{\sum_{k=1}^K e^{z_k}}$$

Then the probability will be:

$$Pr(Y = c | \vec{X} = x) = \frac{e^{w \cdot x + b}}{\sum_{k=1}^K e^{w \cdot x + b}}$$

In Multinomial Logistic Regression, the output variable can have more than two possible discrete outputs. Consider the Digit Dataset.

Ordinal Logistic Regression

It deals with target variables with ordered categories. For example, a test score can be categorized as: “very poor”, “poor”, “good”, or “very good”. Here, each category can be given a score like 0, 1, 2, or 3.

Applying steps in logistic regression modeling:

The following are the steps involved in logistic regression modeling:

- Define the problem: Identify the dependent variable and independent variables and determine if the problem is a binary classification problem.
- Data preparation: Clean and preprocess the data, and make sure the data is suitable for logistic regression modeling.

- Exploratory Data Analysis (EDA)**: Visualize the relationships between the dependent and independent variables, and identify any outliers or anomalies in the data.
- Feature Selection**: Choose the independent variables that have a significant relationship with the dependent variable, and remove any redundant or irrelevant features.
- Model Building**: Train the logistic regression model on the selected independent variables and estimate the coefficients of the model.
- Model Evaluation**: Evaluate the performance of the logistic regression model using appropriate metrics such as **accuracy**, **precision**, **recall**, **F1-score**, or **AUC-ROC**.
- Model improvement**: Based on the results of the evaluation, fine-tune the model by adjusting the independent variables, adding new features, or using regularization techniques to reduce overfitting.
- Model Deployment**: Deploy the logistic regression model in a real-world scenario and make predictions on new data.

Logistic Regression Model Thresholding:

Logistic regression becomes a classification technique only when a decision threshold is brought into the picture. The setting of the threshold value is a very important aspect of Logistic regression and is dependent on the classification problem itself.

The decision for the value of the threshold value is majorly affected by the values of precision and recall. Ideally, we want both precision and recall to be 1, but this seldom is the case.

In the case of a Precision-Recall tradeoff, we use the following arguments to decide upon the threshold:

- 1.Low Precision/High Recall: In applications where we want to reduce the number of false negatives without necessarily reducing the number of false positives, we choose a decision value that has a low value of Precision or a high value of Recall. For example, in a cancer diagnosis application, we do not want any affected patient to be classified as not

affected without giving much heed to if the patient is being wrongfully diagnosed with cancer. This is because the absence of cancer can be detected by further medical diseases but the presence of the disease cannot be detected in an already rejected candidate.

2.High Precision/Low Recall: In applications where we want to reduce the number of false positives without necessarily reducing the number of false negatives, we choose a decision value that has a high value of Precision or a low value of Recall. For example, if we are classifying customers whether they will react positively or negatively to a personalized advertisement, we want to be absolutely sure that the customer will react positively to the advertisement because otherwise, a negative reaction can cause a loss of potential sales from the customer.