

Experiment No. 5

Aim: Implement logistic regression using python to perform classification on social network_ads , cv dataset.

Requirement:

- Anaconda Installer
- Windows 10 OS
- Jupyter Notebook

Theory:

Logistic Regression?

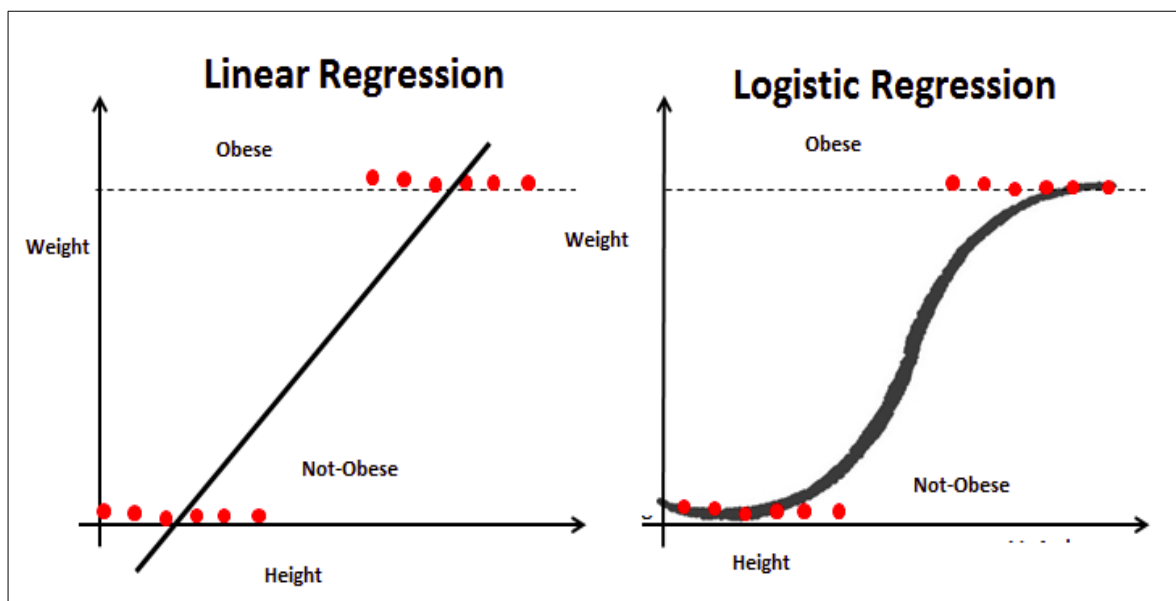


Fig. Linear Regression Vs Logistic Regression

- Logistic regression is a supervised learning algorithm used to predict a dependent categorical target variable. In essence, if you have a large set of data that you want to categorize, logistic regression may be able to help.
- For example, if you were given a dog and an orange and you wanted to find out whether each of these items was an animal or not, the desired result would be for the dog to end up classified as an animal, and for the orange to be categorized as not an animal.
- Animal is your target; it is dependent on your data in order to be able to classify the item correctly. In this example, there are only two possible answers (binary logistic regression), animal or not an animal. However, it is also possible to set up your logistic regression with more than two possible categories (multinomial logistic regression).
- To dive a little deeper into how your model might attempt to classify these two items directly, let's consider what else the model would need to know about the items in order to decide where they belong. Other similar aspects of these items

would need to be looked at when considering how to classify each item or data point. Aspects, or features, may include color, size, weight, shape, height, volume or amount of limbs.

- In this way, knowing that an orange's shape was a circle may help the algorithm to conclude that the orange was not an animal. Similarly, knowing that the orange had zero limbs would help as well.
- Logistic regression requires that the dependent variable, in this case whether the item was an animal or not, be categorical. The outcome is either animal or not an animal—there is no range in between.
- A problem that has a continuous outcome, such as predicting the grade of a student or the fuel tank range of a car, is not a good candidate to use logistic regression. Other options like linear regression may be more appropriate.

Types of Logistic Regression:

There are three main types of logistic regression:

- 1) binary
- 2) multinomial
- 3) ordinal.

They differ in execution and theory. Binary regression deals with two possible values, essentially: yes or no. Multinomial logistic regression deals with three or more values. And ordinal logistic regression deals with three or more classes in a predetermined order.

1. Binary logistic regression:

Binary logistic regression was mentioned earlier in the case of classifying an object as an animal or not an animal—it's an either/or solution. There are just two possible outcome answers. This concept is typically represented as a 0 or a 1 in coding.

Examples include:

- Whether or not to lend to a bank customer (outcomes are yes or no).
- Assessing cancer risk (outcomes are high or low).
- Will a team win tomorrow's game (outcomes are yes or no).

2. Multinomial logistic regression:

Multinomial logistic regression is a model where there are multiple classes that an item can be classified as. There is a set of three or more predefined classes set up prior to running the model.

Examples include:

- Classifying texts into what language they come from.
- Predicting whether a student will go to college, trade school or into the workforce.
- Does your cat prefer wet food, dry food or human food?

3. Ordinal logistic regression:

Ordinal logistic regression is also a model where there are multiple classes that an item can be classified as; however, in this case an ordering of classes is required. Classes do not need to be proportionate. The distance between each class can vary. Examples include:

- Ranking restaurants on a scale of 0 to 5 stars.
- Predicting the podium results of an Olympic event.
- Assessing a choice of candidates, specifically in places that institute ranked-choice voting.

4. Logistic regression assumptions:

- Remove highly correlated inputs.
- Consider removing outliers in your training set because logistic regression will not give significant weight to them during its calculations.
- Does not favor sparse (consisting of a lot of zero values) data.
- Logistic regression is a classification model, unlike linear regression.

Libraries Used:

1. Pandas: Pandas is a Python library used for working with data sets. It has functions for analyzing, cleaning, exploring and manipulating data.
2. Sklearn: It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python.
3. Seaborn: Seaborn is a data visualization library built on top of matplotlib and closely integrated with pandas data structures in Python.
4. Matplotlib: Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib makes easy things easy and hard things possible.

Conclusion:

In this experiment we have studied about the logistic regression model. We ahve performed the classification on the social network. Ads dataset using various liatories of python.