EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Mathematisch-
Naturwissenschaftliche
Fakultät

Logistic Regression

Eberhard Karls Universität Tübingen
Mathematisch-Naturwissenschaftliche Fakultät
Mathias Neitzel

Contents

1 Classification

Classification is a process in which each object of a set (that we give as an input) gets a class assigned/labelled to it.

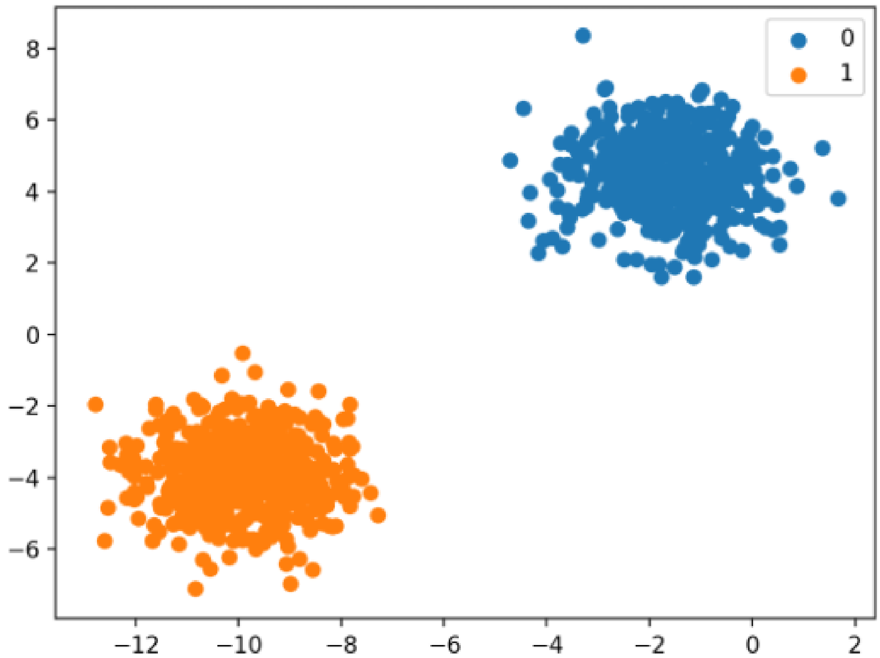
(We will see why this definition is so short.)

1.1 Three Kinds of Classification

There are more kinds of classification depending on how you define/separate them but these three technically include all of them.

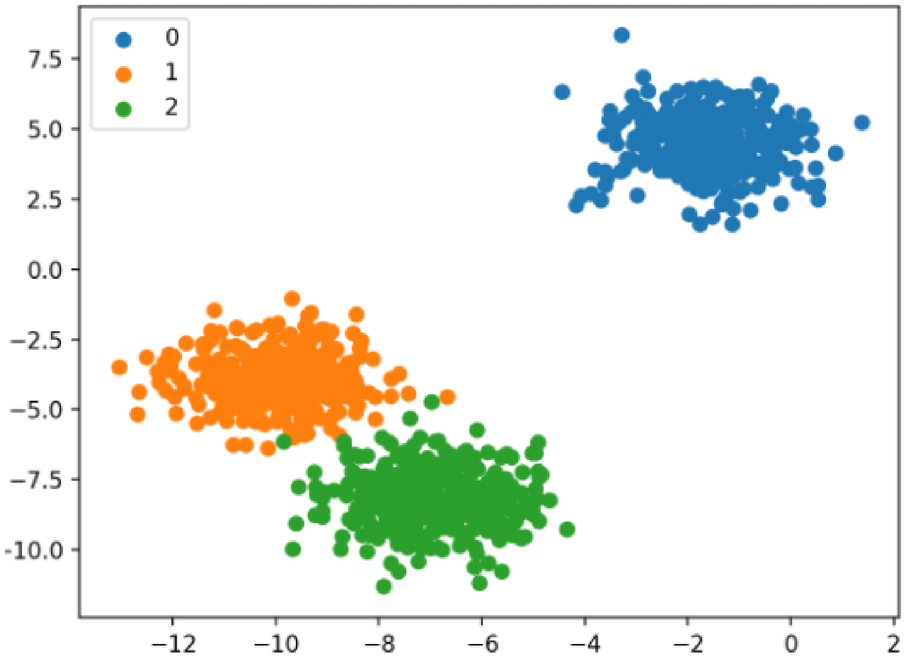
1.1.1 Binary Classification

An object of a set gets labelled into one class while there are exactly two classes into which objects can get labelled.



1.1.2 Multi-Class Classification

An object gets labelled into one class while there are three or more classes into which objects can get labelled.



1.1.3 Multi-Label Classification

An object gets labelled into one or more classes while there are two or more classes into which objects can get labelled.

Multi-Label Classification is why I was so "economical" with the definition of classification.
Often times (also in this seminar) classification is only defined by giving one input exactly one class which would make Multi-Label Classification impossible. So putting more detail into the definition of Classification may would lead to problems.

Multi-Label Classification also sets classification strongly apart from clustering as any clustering can be seen as a classification but not every classification(especially Multi-Label Classification) is a clustering.

To give an example for Multi-Label Classification I would revisit the zoo which we discovered in Hannah Van Santvlients seminar on Unsupervised Learning.

Lets say we have the set of residents of the zoo and the three kinds of food available in the zoo give us our classes.

Residents: chimpanzee, pig, giraffe, fish, banana tree, venus flytrap

Classes: Meat, Plants, Sunlight

When I use Multi-Label Classification we get.

Chimpanzee: Meat, Plants
Pig: Meat, Plants
Giraffe: Plants
Fish: Meat, Plants
Banana Tree: Sunlight
Venus Flytrap: Meat, Sunlight

We can clearly see and understand why we need Multi-Label Classification as the omnivores(or our little exotic friend the venus flytrap) of the zoo would have a hard time by getting only labelled into one Class.

1.1.4 This Seminar

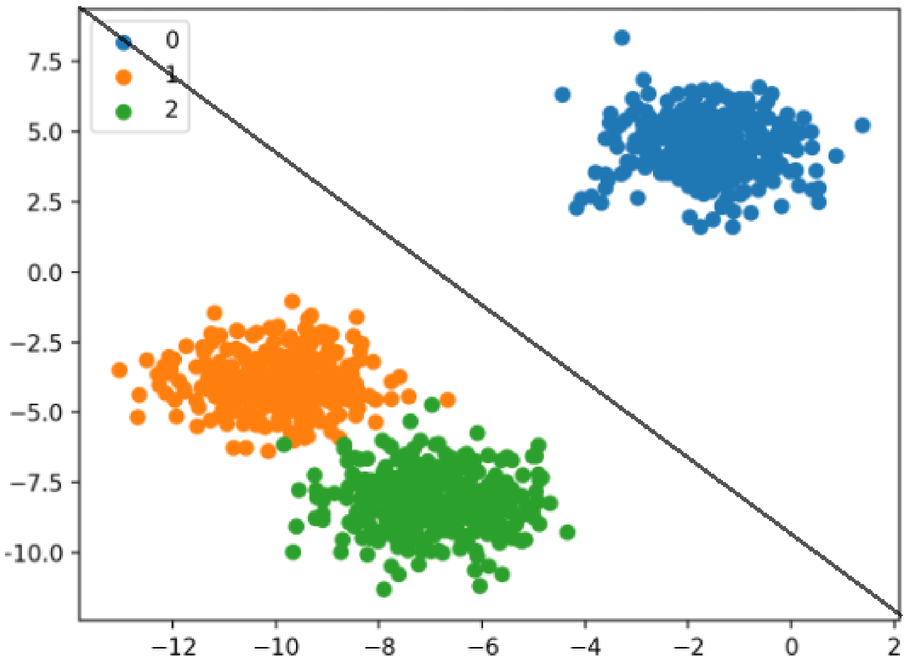
For this seminar I will focus on Binary Classification as the main focus lies on binomial/binary regression.

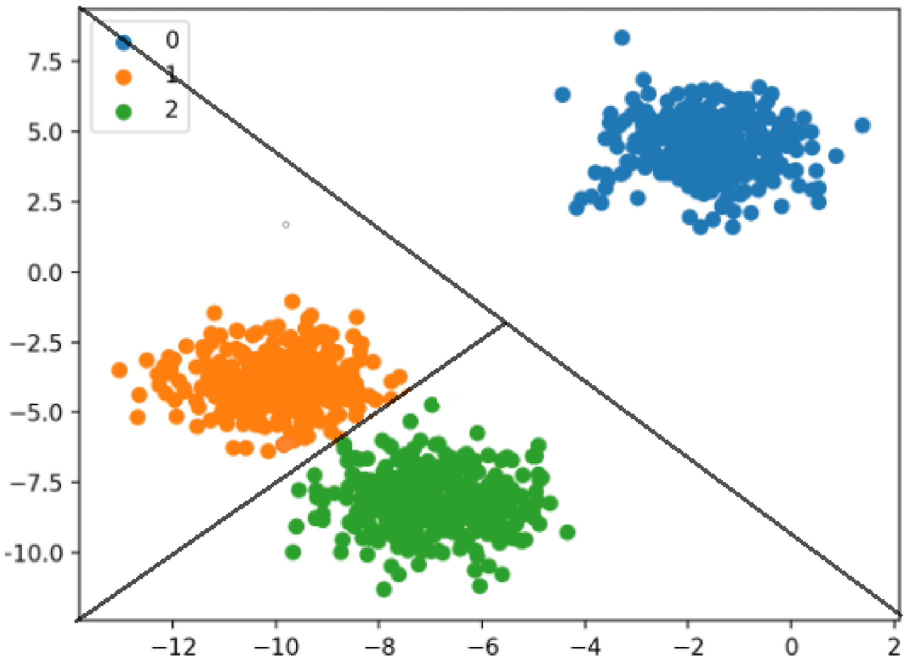
1.2 Linear Models for Classification

In Nadia Vohwinkels talk we learned that we can do a regression trough the Linear Model by finding a function that fits the given data points and enables us to make predictions for possible new points.
The next question one may ask if we can also use the Linear Model for Classification. Martin Kirchhoff will answer this question in detail in his seminar but I have to anticipate a little bit and introduce the topic.

1.2.1 Linear Discriminant Function

A Discriminant Function (or Seperate Function) works by splitting(or seperating) our data set into distinctive groups.





2 Logistic Regression

2.1 Definition of Logistic Regression

Logistic regression uses a logistic function to give a set of data points discrete dependent variables.
When we talk about Multinomial Logistic Regression we can have 2 or more variables but the most used logistic regression is the binary logistic regression where we have exactly two variables. In this seminar I will only talk about binary logistic regression and therefore drop the specification binary.

2.2 The Math of Logistic Regression

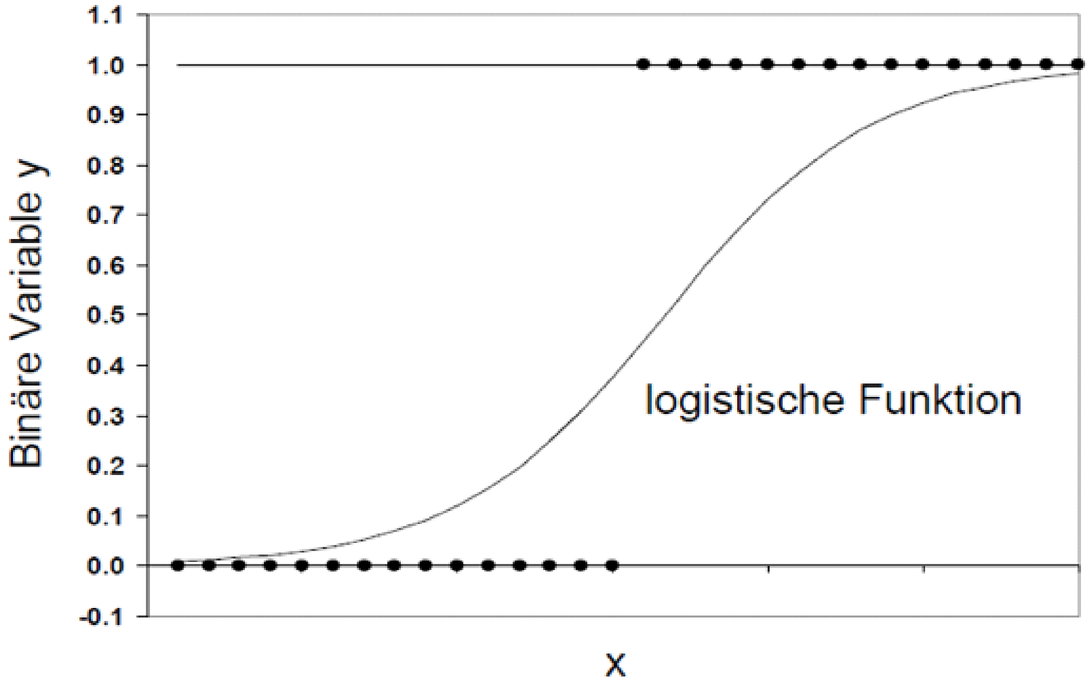
When we use logistic regression we need a set consisting of the fixed coefficients (weights) $x_{i1}x_{i2}\ldots x_{in}$ and one output variable Y.

$(y_i;x_{i1},x_{i2},...,x_{in}),i = 1,...,n$

β is our input: $\beta = (\beta_1,\beta_2,...,\beta_k)$
with these we can finally define our Function as:

$$P(Y_i = 1) = \frac{1}{1+exp(-x_i^T\beta)}$$

This function visualized looks like this:



2.3 Understanding the Sigmoid Function

The sigmoid function itself predicts probabilities this leads to output values in the closed intervall [0,1] one may ask why I talked about the values being discrete or binary. The reason for this is that values of the sigmoid function between [0,0.5) can be projected to 0 and values in (0.5 ,1] can be projected onto the 1 which leads to the values of the logistic regression being binary again.
Values of y_i around 0.5 are ambiguous as it means beta I cant be labelled as 0 or 1 so clearly anymore.
Therefore we can determine how usefull our linear regression is depending on the distribution of the sigmoid function.

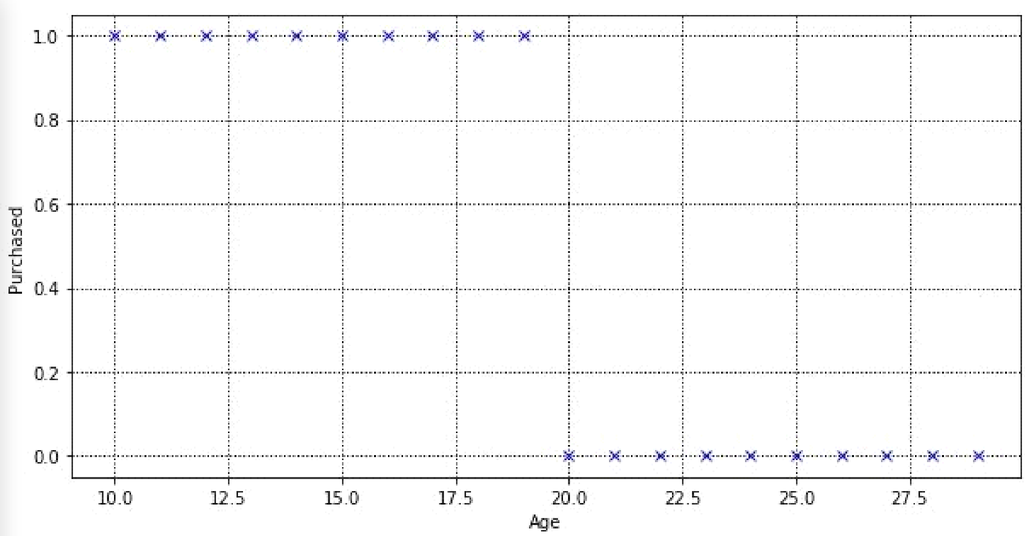
2.4 How do we get the betas/weights?

To get the correct weights for our logistic regression we must train the algorithm with a our available training data. This training is done by using the Maximum-likelihood estimation. The Maximum-likelihood estimation is not only used in logistic regression. We already experienced it in Nadia Vohwinkels seminar on linear regression.

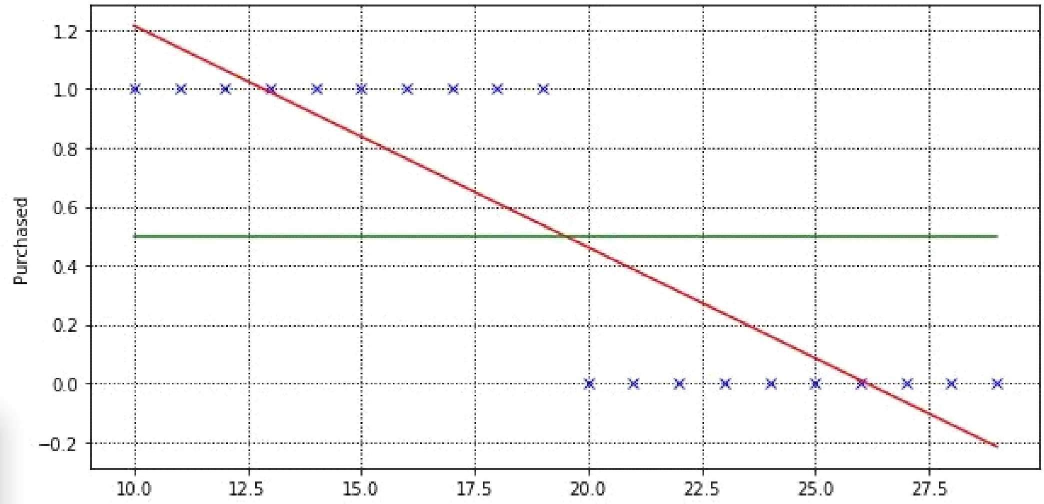
2.5 When to use logistic instead of linear regression?

This can be answered quiet easy. Logistic regression is a model for classification rather than regression even though the name would tell you otherwise. We examine this on an easy example.

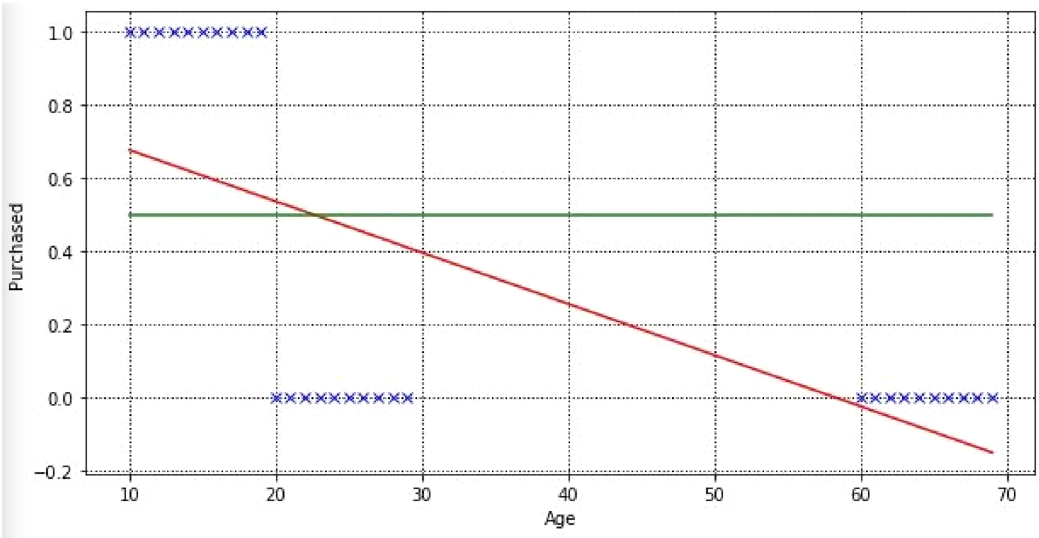
Lets say we have an perfectly balanced dataset of the costumers of a store where everyone is labelled based on if he bought something 1 or he did not 0 and the only input variable is the age.



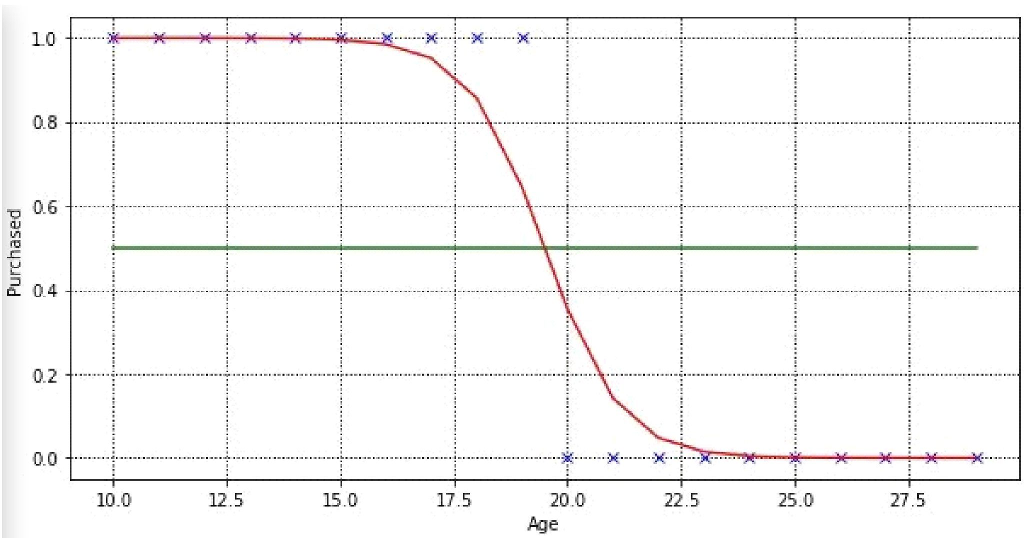
If we let a linear regression algorithm attack that dataset we would get something like that.



As the linear regression algorithm would try to give any input variable(age) an continuous value it would end with values that would not be representative of the buying decision at all. The regression would get even worse when we include older age groups into the dataset.

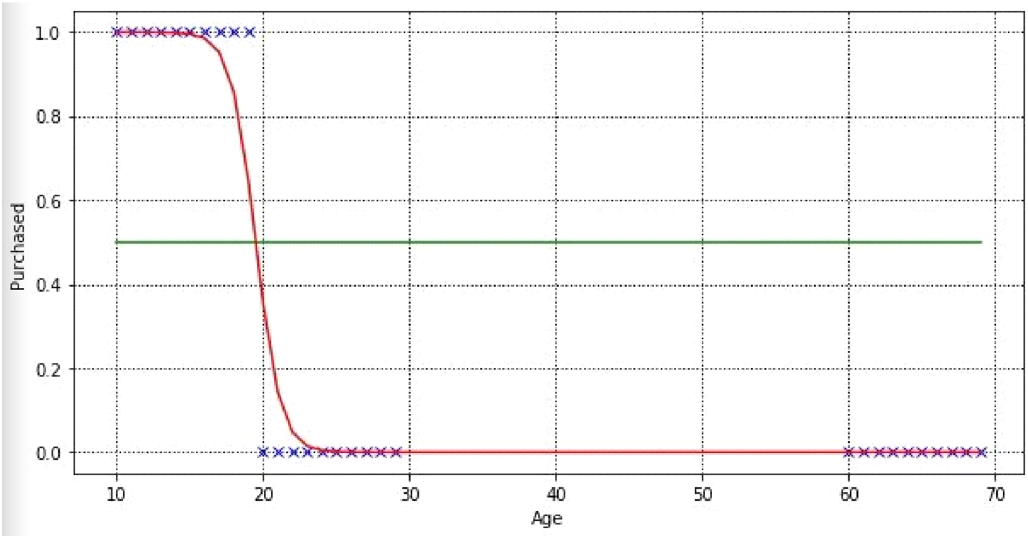


Now we try if a logistic regression algorithm could give use more precise values on the buying decision.



We see that the sigmoid function fits way better into the dataset and is only unprecise in the area around 20 where a split happens in the buying decision.
The same simoid function would also be representative of the dataset which includes

the even older age group.



What we learn out of this example is that in general logistic regression algorithms work better with discrete values as the property of linear regression algorithm to give continuous values hinders there work with discrete values.

3 Quellen

Christopher M. Bishop. Pattern Recognition and Machine Learning. Springer, 2006.

Friedman, Jerome, Trevor Hastie and Robert Tibshirani. The elements of statistical learning. https://christophm.github.io/interpretable-ml-book/logistic.html

Why Linear Regression is not suitable for Classification (the costumer example)

https://jinglescode.github.io/2019/05/07/why-linear-regression-is-not-suitable-for-classification/

Logistic regression
Logistic regression - Wikipedia

Logistische Regression (the german entry heaviely differs from the english one)
Logistische Regression – Wikipedia

Logistic Regression for Machine Learning by Jason Brownlee
Logistic Regression for Machine Learning
(https://machinelearningmastery.com/logistic-regression-for-machine-learning/)

4 Types of Classification Tasks in Machine Learning

4 Types of Classification Tasks in Machine Learning
(https://machinelearningmastery.com/types-of-classification-in-machine-learning/)