Machine Learning

Logistic Regression

Dr. Sherif Saad

Learning Objectives

Introduce a new classification algorithm

Understand how logistic regression works

Understand how to calculate the prediction and how to estimate the weights for a logistic regression model

Logistic Regression

Logistic regression is a classification algorithm not a regression one. We use it to predict class or label not a continuous value.

Logistic regression is the most common applied ML for binary classification problems. Logistic regression is an eager learning method.

Logistic regression is a parametric model where we have m+1 parameters to define the logistic regression model. Here m is the number of features, note that the number of these parameters does not increase or decrease based on the number of training data.

Logistic Regression

Why we call it logistic regression?

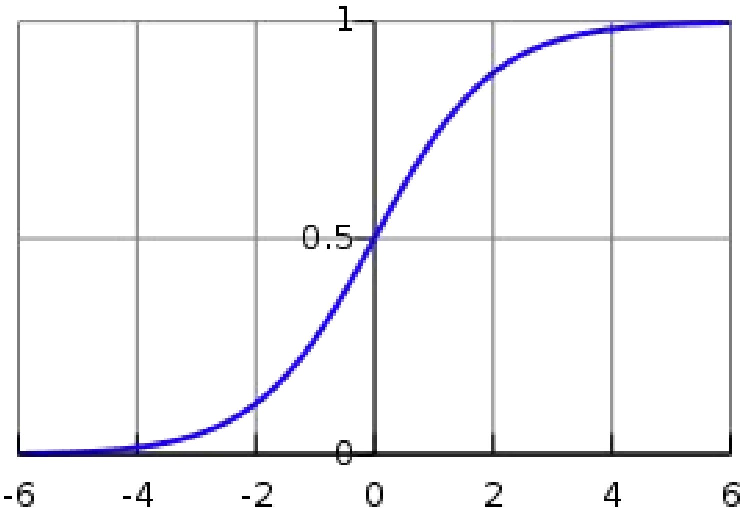
- It is called logistic regression because the core function we use in this model is the logistic function.
- The logistic function also known as the sigmoid function
- It is an S-shaped curve that can take any real value and map it into a value between 0 and 1.

If the value is always between 0 and 1 how we can use it for binary classification?

Logistic Function

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

- ***e*** is the base of the natural logarithms (Euler's number)
- ***x*** is the actual value that we want to transform (map it between 0 and 1)



How can we use logistic function for learning classification?

Logistic Regression Model

The logistic regression model takes a an input value and makes a prediction as to the probability of the input belongs to the class 0 or 1.

If the probability is greater than a predefined threshold we consider the sample belongs to class 1 and if not we consider it belongs to class 0

$$P(Y=1|X) = \frac{1}{1+e^{(w_0+\sum_{i=1}^{n} w_i x_i)}}$$

$$P(Y=0|X) = \frac{e^{(w_0+\sum_{i=1}^{n} w_i x_i)}}{1+e^{(w_0+\sum_{i=1}^{n} w_i x_i)}}$$

Logistic Regression Model

The model is very simple it consist of the logistic functions of given number of parameters and the weights associated with these parameters. At least each feature xi in your data will be associate with w.

The learning process amis at finding or estimating the weights associated with these parameters (features)

The outcome of the learning process is simply the equation and the learned weights. This can be stored into file or in memory.

Learning with Logistic Regression

The weights are estimated from the training data this is usually done using the maximum-likelihood estimation.

The idea is that we want to estimate the weights that would result in a model that predict a value very close to 1 (default class | positive class) and a value very close to 0 for the other class.

The idea behind maximum-likelihood is that we want to minimize the error in the probabilities predicted by logistic regression.

We use a optimization algorithm to find the best weight to minimize the error one simple algorithm to do that is the gradient descent algorithm

Data Preprocessing for Logistic Regression

There are a set of assumptions made by the logistic regression about the data with respect to the data distribution and the relation between the features.

- Binary Classification: it predicts the probability that a sample belongs to one of two available classes.
- Remove Noise: outliers, missing feature values, misclassified samples.
- Remove Correlated Features: highly correlated and dependable features will mostly result in overfitting.
- Gaussian Distribution: performs better when the features has a normal distribution.
- Feature Linearity: it assume linearity of the independent variables.

Making Prediction with Logistic Regression

Let us assume we are using the breast cancer.

We only have two features these features are Tumor Radius and Tumor Size.

We have 12 samples in our dataset and all the samples are labeled

ID	Tumor Radius	Tumor Size	Class
1	1.46	2.8	B
2	7.7	2.8	M
3	2.7	2.4	B
4	5.4	2.1	M
5	4.1	3.7	M
6	3.3	4.4	B
7	6.9	2.7	M
8	3.1	3.2	B
9	8.5	3.7	M
10	1.4	2.3	B
11	7.8	3.4	M
12	4.1	3.7	B

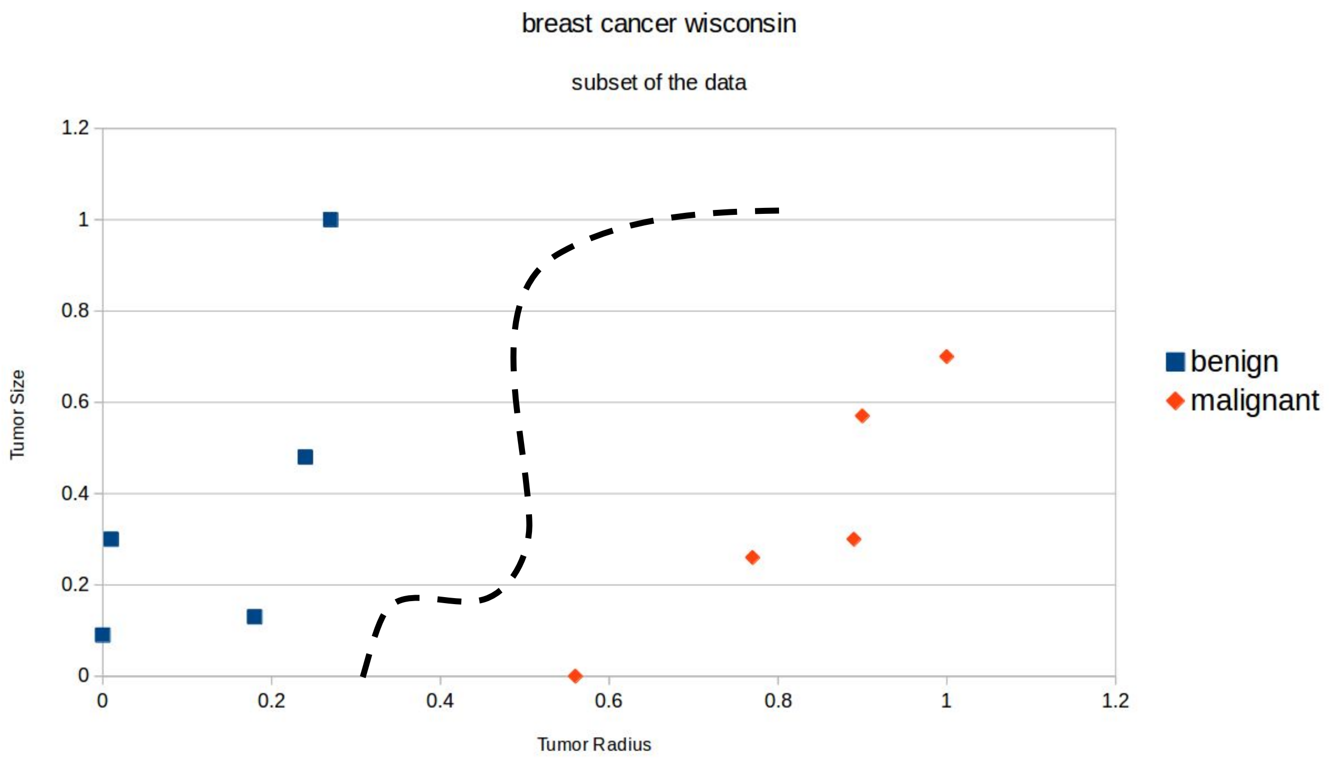
Making Prediction with Logistic Regression

$$new(x_{i,j}) = \frac{x_{i,j} - min(X_{i,j})}{max(X_{i,j}) - min(X_{i,j})}$$

ID	Tumor Radius	Tumor Size	Class
1	1.46	2.8	B
2	7.7	2.8	M
3	2.7	2.4	B
4	5.4	2.1	M
5	4.1	3.7	M
6	3.3	4.4	B
7	6.9	2.7	M
8	3.1	3.2	B
9	8.5	3.7	M
10	1.4	2.3	B
11	7.8	3.4	M
12	4.1	3.7	B

ID	Tumor Radius	Tumor Size	Class
1	0.01	0.30	0
2	0.89	0.30	1
3	0.18	0.13	0
4	0.56	0.00	1
5	0.38	0.70	1
6	0.27	1.00	0
7	0.77	0.26	1
8	0.24	0.48	0
9	1.00	0.70	1
10	0.00	0.09	0
11	0.90	0.57	1
12	0.38	0.70	0

Making Prediction with Logistic Regression



Making Prediction with Logistic Regression

For this subset of the data the logistic regression has 3 weights to estimate (learn) so we can distinguish between benign and malignant tumors.

The expected logistic regression model will look like

$$LR(x) = g(w_0 + w_1 x_1 + w_2 x_2)$$

The learning process aims at finding the best value for weights with respect to the training data.

Making Prediction with Logistic Regression

How can we estimate the weights for logistic regression?

- There are several methods to do that, here we will use stochastic gradient descent algorithm.

The Stochastic Gradient Descent algorithm is a very simple and powerful method that we can use in different machine learning models.

The idea of SGD is simple we use the ML model to calculate the prediction for each instance| sample in the training dataset and calculate the error for each prediction.

Stochastic Gradient Descent

To find the weights for a logistic regression model we:

1. Initialize the weights using some random values.
2. For each sample calculate the prediction using the current weights
3. Calculate the prediction error and use it to update the weights
4. Repeat step 2 and 3 until the model reach the desired accuracy or for a fixed number of iteration.

Here we are using online learning model since we update the weight after each prediction. Another approach is batch learning where we use all the recorded errors and perform all the update in one step.

Predication and Weight Estimation

Given the subset of the breast cancer dataset. We will show the main steps to construct the logistic regression model (learning the weight)

Let us initialize all the 3 weights by the value zero. Therefore at the beginning w0, w1 and w2 =0.

Now let us use the first instance and calculate the prediction and update the weights based on the prediction error. Here we are using online line learning.

In your opinion what is better using online learning or batch learning?

Predication and Weight Estimation

Using the first sample and the deafult weight we calculate the prediction as follows

$$P(X_1) = \frac{1}{1 + e^{-(w_0 + w_1 x_{1,1} + w_2 x_{1,2})}}$$

$$P(X_{i,j}) = \frac{1}{1 + e^{-(0+0*0.01+0*0.3)}} = 0.5$$

ID	Tumor Radius	Tumor Size	Class
1	0.01	0.30	0
2	0.89	0.30	1
3	0.18	0.13	0
4	0.56	0.00	1
5	0.38	0.70	1
6	0.27	1.00	0
7	0.77	0.26	1
8	0.24	0.48	0
9	1.00	0.70	1
10	0.00	0.09	0
11	0.90	0.57	1
12	0.38	0.70	0

Predication and Weight Estimation

After calculating the prediction we calculate the prediction error and update the weights. Using the following equation

$$new(w_i) = w_i + \alpha \cdot (l - p) \cdot p \cdot (1 - p) \cdot x_i$$

Where w_i is the weight we want to update, alpha is the learning rate (preselected value) that controls how the weights are updated. L is the expected label and p is the predicate label and x_i is the feature associated with w_i

ID	Tumor Radius	Tumor Size	Class
1	0.01	0.30	0
2	0.89	0.30	1
3	0.18	0.13	0
4	0.56	0.00	1
5	0.38	0.70	1
6	0.27	1.00	0
7	0.77	0.26	1
8	0.24	0.48	0
9	1.00	0.70	1
10	0.00	0.09	0
11	0.90	0.57	1
12	0.38	0.70	0

Predication and Weight Estimation

$$new(w_0) = 0 + 0.7 \cdot (0 - 0.5) \cdot 0.5 \cdot (1 - 0.5) \cdot 1 = -0.08$$

$$new(w_1) = 0 + 0.7 \cdot (0 - 0.5) \cdot 0.5 \cdot (1 - 0.5) \cdot 0.01 = -0.0008$$

$$new(w_1) = 0 + 0.7 \cdot (0 - 0.5) \cdot 0.5 \cdot (1 - 0.5) \cdot 0.3 = -0.026$$

How do we know that the model is learning?

ID	Tumor Radius	Tumor Size	Class
1	0.01	0.30	0
2	0.89	0.30	1
3	0.18	0.13	0
4	0.56	0.00	1
5	0.38	0.70	1
6	0.27	1.00	0
7	0.77	0.26	1
8	0.24	0.48	0
9	1.00	0.70	1
10	0.00	0.09	0
11	0.90	0.57	1
12	0.38	0.70	0

Predication and Weight Estimation

$$new(w_0) = 0 + 0.7 \cdot (0 - 0.5) \cdot 0.5 \cdot (1 - 0.5) \cdot 1 = -0.08$$

$$new(w_1) = 0 + 0.7 \cdot (0 - 0.5) \cdot 0.5 \cdot (1 - 0.5) \cdot 0.01 = -0.0008$$

$$new(w_1) = 0 + 0.7 \cdot (0 - 0.5) \cdot 0.5 \cdot (1 - 0.5) \cdot 0.3 = -0.026$$

How do we know that the model is learning?

$$P(X_{i,j}) = \frac{1}{1 + e^{-(-0.08 - 0*0008*0.01 - 0.026*0.3)}} = 0.47$$

ID	Tumor Radius	Tumor Size	Class
1	0.01	0.30	0
2	0.89	0.30	1
3	0.18	0.13	0
4	0.56	0.00	1
5	0.38	0.70	1
6	0.27	1.00	0
7	0.77	0.26	1
8	0.24	0.48	0
9	1.00	0.70	1
10	0.00	0.09	0
11	0.90	0.57	1
12	0.38	0.70	0

Predication and Weight Estimation

We repeat the process for of predication and weight estimation for n number of iterations. In each iteration we update the weights by going over all the instances in the training dataset.

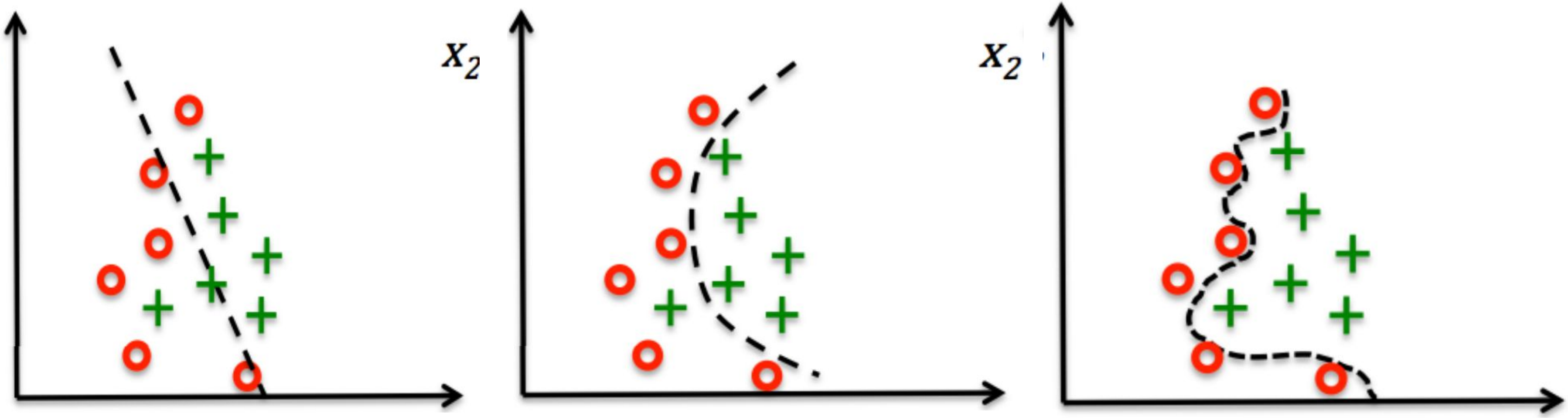
Each iteration that goes over the training dataset is called an epoch

We can either terminate after n number of iterations or if the accuracy or the error rate are not improving.

When do you think the accuracy or the error rate will not improve?

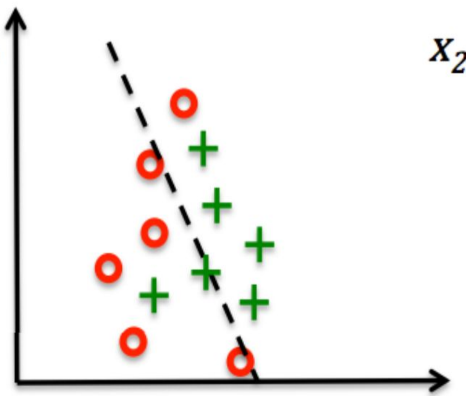
Logistic Regression and Overfitting

When does overfitting happen in logistic regression, is it even possible?

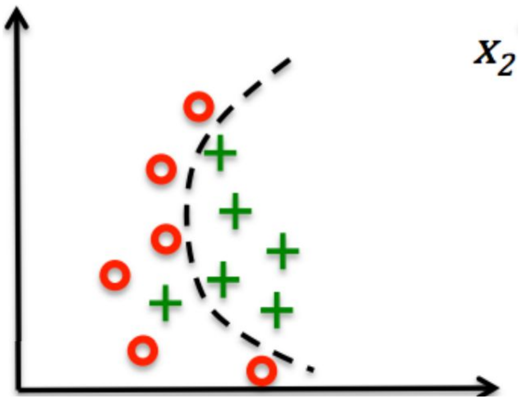


Logistic Regression and Overfitting

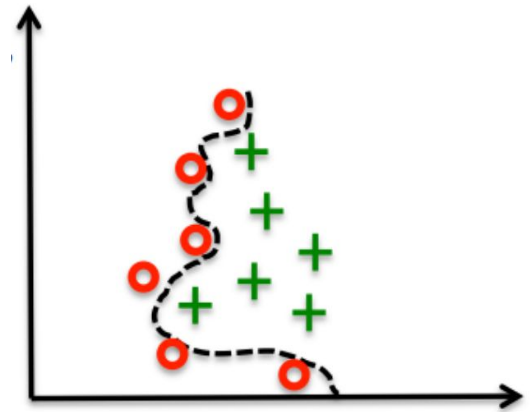
When does overfitting happen in logistic regression, is it even possible?



$$LR(x) = g(w_0 + w_1 x_1 + w_2 x_2)$$



$$LR(x) = g(w_0 + w_1 x_1 + w_2 x_2 + w_3 x_1^2 + w_4 x_2^2 + w_5 x_1 x_2)$$



$$LR(x) = g(w_0 + w_1 x_1 + w_2 x_2 + w_3 x_1^2 + w_4 x_2^2 + w_5 x_1 x_2 + w_6 x_1^2 x_2^2 + w_6 x_1^2 x_2^3 + w_6 x_1^3 x_2^2 + \ldots)$$

Questions