Linear Regression using Machine Learning

Yi Qiang Ji Zhang
Biel Galiot Pérez
Cristian Asensio Garcia
Professor: Alex Ferrer Ferre

Aerospace Engineering
Numerical Tools in Machine Learning for Aeronautical Engineering
Polytechnic University of Catalonia — BarcelonaTech

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Escola Superior d'Enginyeries Industrial, Aeroespacial i Audiovisual de Terrassa

Date: 24 May 2021

I. Introduction

The classification problem is similar to the regression issue, with the exception that the values we wish to predict now have just a few discrete values. For the time being, we'll review the binary classification issue, where y has only two possible values: 0 and 1 to introduce the hypothesis representation.

Hypothesis Representation

The essence of cost function and advanced optimization techniques Unlike linear regression, the hypothesis function of logistic regressions includes a sigmoid function:

$$\begin{gathered} h_\theta(x) = g\left(\theta^T x\right) \\ z = \theta^T x \\ g(z) = \frac{1}{1+e^{-z}} \\ \\ h_\theta(x) = P(y=1 \mid x;\theta), \\ P(y=0 \mid x;\theta) + P(y=1 \mid x;\theta) = 1 \end{gathered} \tag{1}$$

In which the hypothesis will generate a probability that $y = 1$ (or $y = 0$) given input x, with the threshold set to a certain number between 0 and 1 . For example, we can predict $y = 1$ if $h(x) => 0.5$ or 0.7, and y = 0 if h(x) < 0.5 or 0.3 respectively.

The function $g(z)$, shown here, maps any real number to the $(0, 1)$ interval, making it useful for transforming an arbitrary-valued function into a function better suited for classification.

The input to the sigmoid function $g(z)$ (e.g. $\theta^T X$) doesn't need to be linear, and could be a function that describes a circle (e.g. $z = \theta_0 + \theta_1 x_1^2 + \theta_2 x_2^2$) or any shape to fit our data.

Cost function:

$$\begin{aligned} J(\theta) = \frac{1}{m}\sum_{i=1}^{m} \mathrm{Cost}\left(h_\theta\left(x^{(i)}\right), y^{(i)}\right) = \\ -\frac{1}{m}\left[\sum_{i=1}^{m} y^{(i)} \log\left(h_\theta\left(x^{(i)}\right)\right) + \right. \\ \left. \left(1-y^{(i)}\right)\log\left(1-h_\theta\left(x^{(i)}\right)\right)\right] \end{aligned} \tag{2}$$

II. Linear Regression XClassification

Iris data set

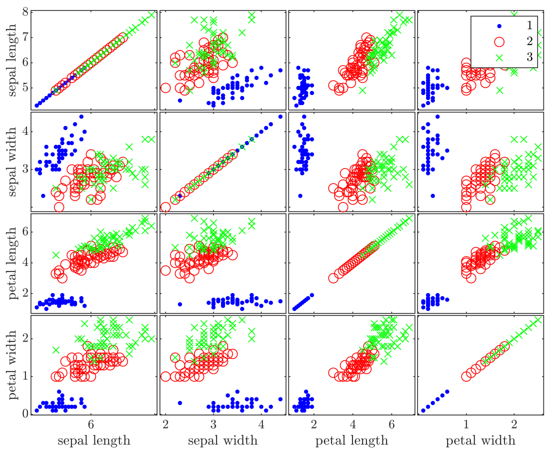

Fig. 1. Iris set plot.

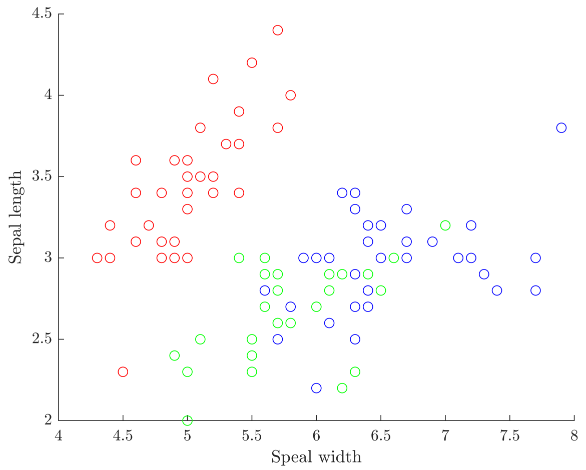

Fig. 2. Sepal Width vs Sepal length.