Exercise 4: Part 3

Linear Regression

Python Programming Bootcamp by Dr Rohitash Chandra
UNSW, 2021

Introduction

Your are required to use ***object oriented programming*** techniques in your program.

Description

Iris classification data set is a well-known data set in machine learning. More details of the data set can be found here. Iris flower data set. (2020). Wikipedia. Retrieved from https://en.wikipedia.org/wiki/Iris_flower_data_setm

"The data set consists of 50 samples from each of three species of Iris (Iris setosa, Iris virginica and Iris versicolor). Four features were measured from each sample: the length and the width of the sepals and petals, in centimeters. Based on the combination of these four features, Fisher developed a linear discriminant model to distinguish the species from each other."

1. Compute the R squared score among feature 1 (column 1) and feature 3 (column 3) from original Iris dataset after developing a linear regression model. Carry out the same for feature 1 vs feature 2 and feature 1 vs feature 4. What do you observe and what can you conclude?
2. Try developing a linear regression model for binary classification using the step transfer function using the data produced from Iris binary classification case. Use 60 percent of data picked randomly for training and remaining for testing. Try the approach with all 4 input features, i) without normalising input data, ii) with normalising input data. Report percentage correctly classified in training and test set after fixed number of training time (iterations) using gradient descent. Carry out 30 experiments for both cases with different initial random seeds for the parameters of concern. Report mean and standard decision using the appropriate metrics for your results. The purpose of the exercise is to brainstorm if this can be done. Step function is given below:

$$f(x) = 1, x>=0 = 0, x<0$$

Resources

https://www.analyticsvidhya.com/blog/2020/01/fundamentals-deep-learning-activation-functions-when-to-use-them/