



ft_linear_regression

An introduction to machine learning

Baptiste Jacob bjacob@student.42.fr

Résumé: In this project you are going to implement your first machine learning algorithm.

Table des matières

I	Préambule	2
II	Introduction	3
III	Objective	4
IV	General instructions	5
V	What you have to do	6
VI	Bonuses	7
VII	Peer-evaluation	8

Chapitre I

Préambule

What i think is the best definition for machine learning :

“A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P , if its performance at tasks in T , as measured by P , improves with experience E ”

Tom M. Mitchell

Chapitre II

Introduction

Machine learning is a growing field of computer science that may seem a bit complicated and only reserved to mathematicians. You may have heard of neural network or k-means clustering and don't understand how they work or how to code this kind of algorithm. Don't worry on this subject we will start with a basic and very simple machine learning algorithm.

Chapitre III

Objective

The objective of this project is to enable you to understand basic concept of machine learning. In this subject you will have to create a program that predicts the price of a car by using a [linear function](#) train with a [gradient descent algorithm](#).

We will be using a precise example in this subject, but at the end you will be able to use this algorithm with any other dataset.

Chapitre IV

General instructions

In this project you are free to use whatever language you want. If you use C, you have to respect the Norm, as usual.

You are also free to use any libraries you want as long as it does not do all the work for you. For example, if you use python `numpy.polyfit` is considered as a cheat case.



You should use a language that allows you to easily visualize your data it will be very helpful when you will have to debug

Chapitre V

What you have to do

In this project you will realize a simple linear regression with a single feature which is the mileage of the car.

To do so you need to create two programs :

- The first program will be use to predict the price of a car for a given mileage. If you launch this program it will prompt you for a mileage and print you the estimated price for this mileage. This program will use the following hypotheses to predict the price :

$$estimatePrice(mileage) = \theta_0 + (\theta_1 * mileage)$$

Before we run our training program θ_0 and θ_1 will be set to 0.

- The second program will be use to train our model. This program will read our dataset file and perform a linear regression in this data. After performing the linear regression you will save your variable θ_0 and θ_1 to use it in your first program. You will be using the following formula :

$$tmp\theta_0 = learningRate * \frac{1}{m} \sum_{i=0}^{m-1} (estimatePrice(mileage[i]) - price[i])$$

$$tmp\theta_1 = learningRate * \frac{1}{m} \sum_{i=0}^{m-1} (estimatePrice(mileage[i]) - price[i]) * milleage[i]$$

I let you guess what m is :)

note that the estimatePrice is the same as in our first program but it use our temporary computed θ_0 and θ_1 . Also don't forget to simultaneously update θ_0 and θ_1 .

Chapitre VI

Bonuses

- plotting the data into a graph to see there repartition.
- plotting the line that fit the data to see the result of our work.
- create a program to calculate the precision of our algorithm.

Any other bonuses that can make your program awesome.

Chapitre VII

Peer-evaluation

The peer who will grade you will have to verify :

- for cheat function
- if hypotheses is the same as specified before
- if taining function is the same as specified before