Theoretical Machine Learning

Week 2 : Linear Regression

18th May 2024

Introduction

Say, you have a relation, or graph like this.

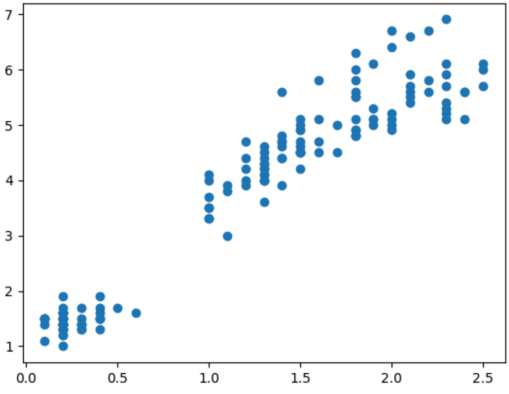

Figure: 1

Now, you want to find the best-fit curve, to approximate y with a given x, or basically regress y on x. How would you solve it mathematically? Looking at the graph, we can draw a line which appears reasonable, but would that be the best-fit line, and how do we define a best-fit line?

Loss Function

Say, we get the function f(X) to predict Y. We would then want to see how close or far off our function is to the actual value of Y, or basically how accurate our function is. More the accuracy, the better our function.

Typically, our loss function looks like,
$L(X, Y) = \frac{\Sigma_i g(|Y_i - f(X_i)|)}{N}$, where g is an increasing function in $\mathbb{R}^+$. Our task is to minimise this.

The most common one which you might have encountered is when g(X) $= X^2$, or the Mean Squared Error. This is also called L2 - Norm.
When we have g(X) $= |X|$, we get L1 - Norm.
You might now ask why would we not just stick to one method, for uniformity.
Again, it depends on the situation. For L1 Norm, you cannot find a closed form solution. In L2 Norm, in a lot of cases, you get a closed form solution. However, the reason we sometimes use L1 Norm is because it does not inflate with crazy outliers.
For this week, we look at the linear case, with an L2 Norm. So, basically, our f is a linear function, and our loss function is a mean squared error.