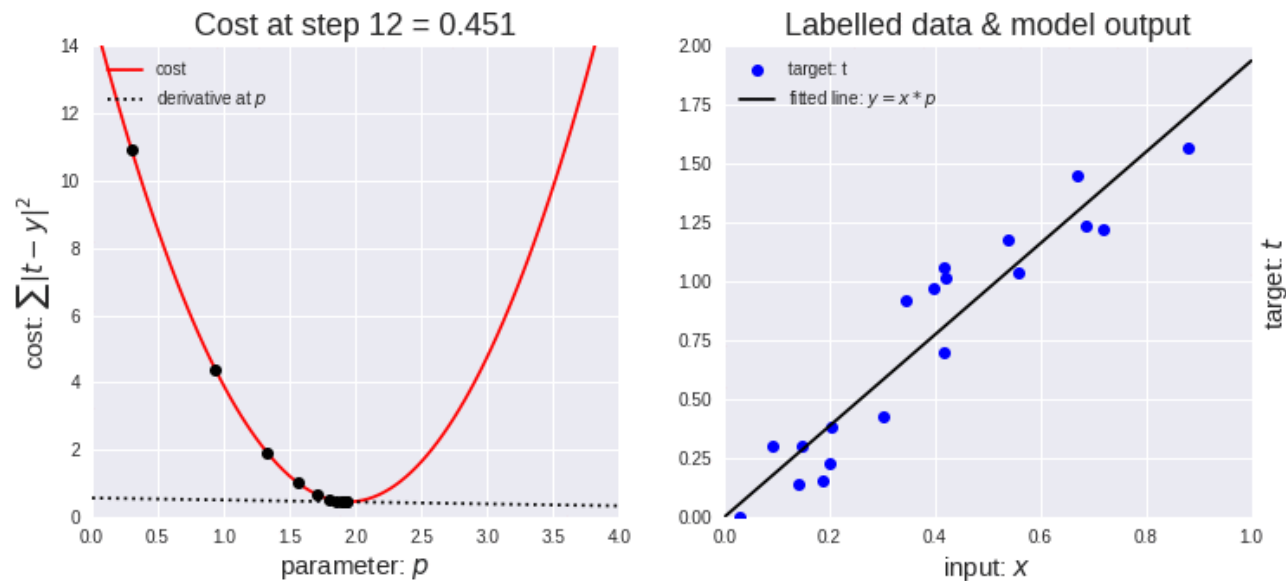


Epoch VS Batch Size VS Iterations

Gradient Descent

It is an *iterative* optimization algorithm used in machine learning to find the best results (minima of a curve).

Gradient means the *rate* of inclination or declination of a slope. *Descent* means the instance of *descending*.



The algorithm is *iterative* means that we need to get the results multiple times to get the most optimal result. The iterative quality of the gradient descent helps a under-fitted graph to make the graph fit optimally to the data.

Epochs

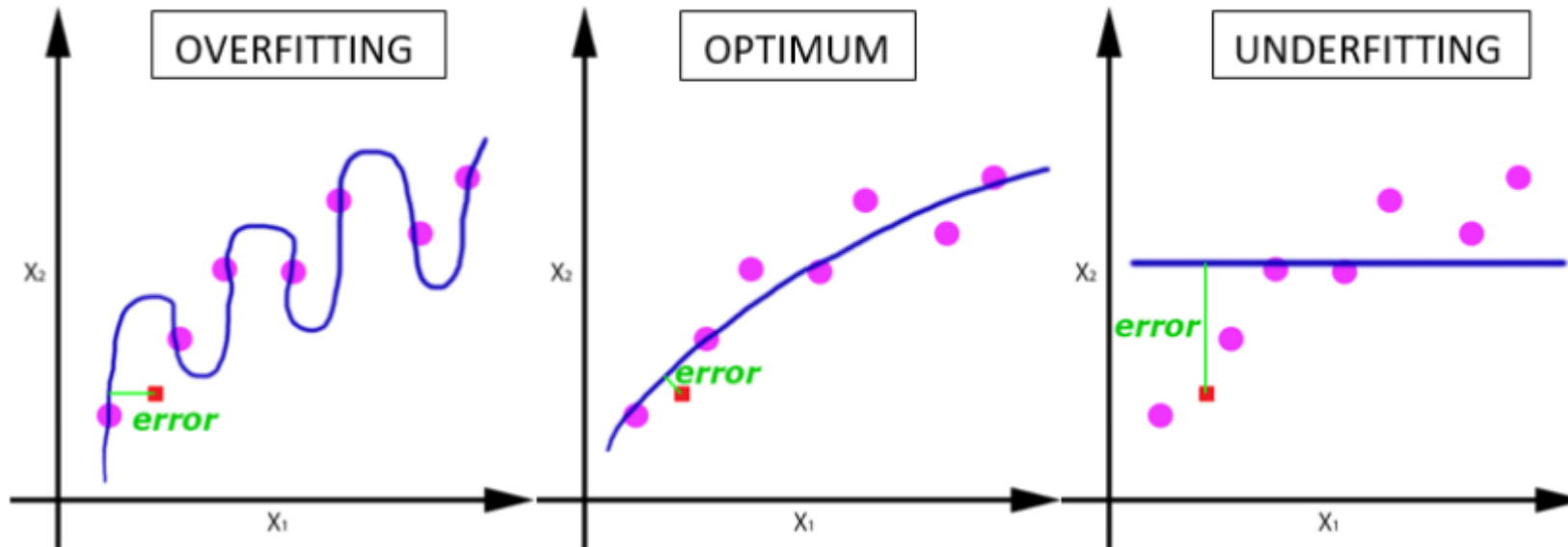
“One Epoch is when an ENTIRE dataset is passed forward and backward through the neural network only ONCE”

Since one epoch is too big to feed to the computer at once we divide it in several smaller batches

Why we use more than one Epoch?

I know it doesn't make sense in the starting that — passing the entire dataset through a neural network is not enough. And we need to pass the full dataset multiple times to the same neural network. But keep in mind that we are using a limited dataset and to optimise the learning and the graph we are using **Gradient Descent** which is an *iterative* process. So, *updating the weights with single pass or one epoch is not enough*.

One epoch leads to under fitting of the curve in the graph (below).



As the number of epochs increases, more number of times the weight are changed in the neural network and the curve goes from **underfitting** to **optimal** to **overfitting** curve.

So, what is the right numbers of epochs?

Unfortunately, there is no right answer to this question. The answer is different for different datasets but you can say that the numbers of epochs is related to how diverse your data is... just an example - Do you have only black cats in your dataset or is it much more diverse dataset?

Batch Size

“Total number of training examples present in a single batch”

Batch size and number of batches are two different things

What is a Batch?

As I said, you can't pass the entire dataset into the neural net at once. So, you **divide dataset into Number of Batches or sets or parts.**

Just like you divide a big article into multiple sets/batches/parts like Introduction, Gradient descent, Epoch, Batch size and Iterations which makes it easy to read the entire article for the reader and understand it. 😊

Iterations

“Iterations is the number of batches needed to complete one epoch.”

The number of batches is equal to number of iterations for one epoch.

Let's say we have 2000 training examples that we are going to use.

We can divide the dataset of 2000 examples into batches of 500 then it will take 4 iterations to complete 1 epoch.

Where Batch size is 500 and iterations is 4, for 1 complete epoch.