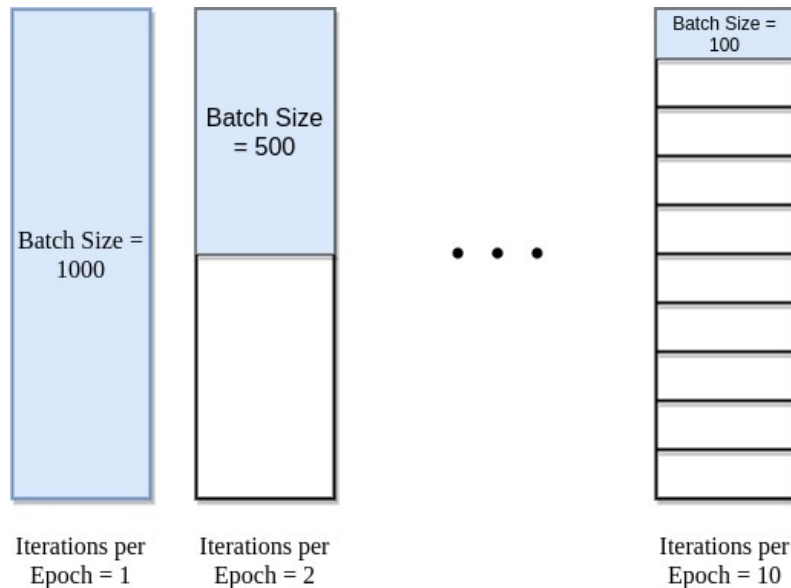An epoch is made up of one or more batches, where we use a part of the dataset to train the neural network. We call passing through the training examples in a batch an iteration.

An epoch is sometimes mixed with an iteration. To clarify the concepts, let's consider a simple example where we have 1000 data points as presented in the figure below:
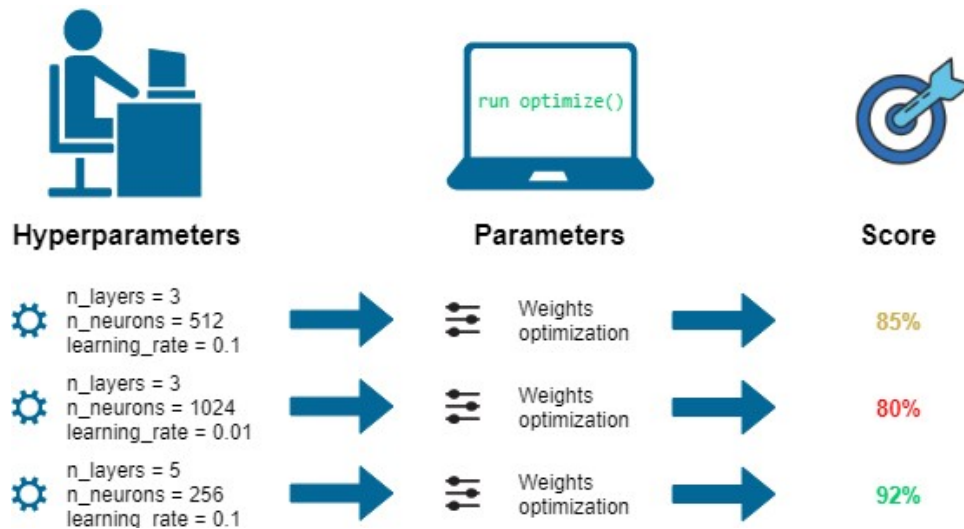

If the batch size is 1000, we can complete an epoch with a single iteration. Similarly, if the batch size is 500, an epoch takes two iterations. So, if the batch size is 100, an epoch takes 10 iterations to complete. Simply, for each epoch, the required number of iterations times the batch size gives the number of data points.

We can use multiple epochs in training. In this case, the neural network is fed the same data more than once.

5. HYPER PARAMETER TUNING

In machine learning, **hyperparameter optimization** or tuning is the problem of choosing a set of optimal hyperparameters for a learning algorithm. A hyperparameter is a parameter whose value is used to control the learning process. By contrast, the values of other parameters (typically node weights) are learned.

The same kind of machine learning model can require different constraints, weights or learning rates to generalize different data patterns. These measures are called hyperparameters, and have to be tuned so that the model can optimally solve the machine learning problem. Hyperparameter optimization finds a tuple of hyperparameters that yields an optimal model which minimizes a predefined loss function on given independent data

Hyperparameters		Parameters		Score
n_layers = 3 n_neurons = 512 learning_rate = 0.1	→	Weights optimization	→	85%
n_layers = 3 n_neurons = 1024 learning_rate = 0.01	→	Weights optimization	→	80%
n_layers = 5 n_neurons = 256 learning rate = 0.1	→	Weights optimization	→	92%

Approaches

Grid search

One traditional and popular way to perform hyperparameter tuning is by using an Exhaustive Grid Search. This method tries every possible combination of each set of hyper-parameters. Using this method, we can find the best set of values in the parameter search space. This usually uses more computational power and takes a long time to run since this method needs to try every combination in the grid size.

Randomized Search

The main difference in the RandomizedSearch CV, when compared with GridCV, is that instead of trying every possible combination, this chooses the hyperparameter sample combinations randomly from grid space. Because of this reason, there is no guarantee that we will find the best result like Grid Search. But, this search can be extremely effective in practice as computational time is very less.