Keras Optimizers

There are many optimizers in Keras, that we encourage you to explore further, in this **link**, or in this excellent **blog post**. These optimizers use a combination of the tricks above, plus a few others. Some of the most common are:

SGD

This is Stochastic Gradient Descent. It uses the following parameters:

- Learning rate.
- Momentum (This takes the weighted average of the previous steps, in order to get a bit of momentum and go over bumps, as a way to not get stuck in local minima).
- Nesterov Momentum (This slows down the gradient when it's close to the solution).

Adam

Adam (Adaptive Moment Estimation) uses a more complicated exponential decay that consists of not just considering the average (first moment), but also the variance (second moment) of the previous steps.

RMSProp

RMSProp (RMS stands for Root Mean Squared Error) decreases the learning rate by dividing it by an exponentially decaying average of squared gradients.

NEXT