**Day 21**

**What to do?**

Learn about Adam optimizer.

**Adam Optimizer:**

Since Adagrad takes an exceedingly long time to converge, it has been modified and optimized to Adam (Adaptive Moment Estimation) optimizer. Adam optimizer is combination of momentum gradient descent and RMSProp optimizer. It implements exponentially moving average of the gradients to scale learning rate and optimize model. It is one of the quickest optimizers in deep neural networks.

The algorithm first implements exponentially moving average of the gradient (vdw and vdb), then implements square gradient (RMSProp, sdw and sdb).

vdw = beta\_1 \* vdw + (1 – beta\_1) \* dW; vdb = beta\_1 \* vdb + (1 – beta\_1) \* db

sdw = beta\_2 \* sdw + (1 – beta\_2) \* dW2; sdb = beta\_2 \* sdb + (1 – beta\_2) \* db

The additional calculation in Adam optimizer is bias correction.

vdwcorrected = vdw / (1 – beta\_1); vdbcorrected = vdb / (1 – beta\_1)

sdwcorrected = sdw / (1 – beta\_2); sdbcorrected = sdb / (1 – beta\_2)

Finally, the weights and the biases are updated, using the following equations.

W = W – alpha \* vdwcorrected / (sdwcorrected + epsilon)1/2

b = b – alpha \* vdbcorrected / (sdbcorrected + epsilon)1/2

In this optimizer, the hyperparameters are alpha, beta\_1, beta\_2 and epsilon. Usually, the hyperparameter value alpha can be hyper tuned, whereas beta\_1 can be 0.9, beta\_2 can be 0.999 and epsilon 10-8.