**Day 20**

**What to do?**

Learn about AdaGrad optimizer.

**AdaGrad Optimizer:**

AdaGrad has slight variation to RMSProp. Let us look at the equations:

vdw = vdw + dw2

vdb = vdb + db2

W = W – alpha \* dw/(vdw1/2 + epsilon)

b = b – alpha \* db/(vdb1/2 + epsilon)

1. *AdaGrad*

vdw = beta \* vdw + (1 – beta) \* dw2

vdb = beta \* vdb + (1 – beta) \* db2

W = W – alpha \* dw/(vdw1/2 + epsilon)

b = b – alpha \* db/(vdb1/2 + epsilon)

1. *RMSProp*

The only difference between AdaGrad and RMSProp is the decay rate “beta”. Moreover, the disadvantage with AdaGrad is the lack of decay rate, which leads to huge implementation time to update the weights (especially when vdw and vdb become large, which they most probably will overtime).