**Day 23**

**What to do?**

Learn which optimizer to choose when.

While there are many optimizers to choose from, it sometimes bugs people (even me) in choosing the right optimizer. I usually just choose Adam optimizer for every type of problem and dataset. However, there is an entire intuition on when to use the optimizers that justifies its purpose.

1. The disadvantage of AdaGrad optimizer is it penalizes learning rate too much when the parameters update too frequently. Hence, AdaGrad works best when the features are sparse. In other words, they should not update too often.
2. RMSProp optimizer works on the disadvantage of AdaGrad optimizer. The data does not need to be sparse as it can update parameters more frequently with no penalty.
3. Adam optimizer (my personal favorite) is better than the other optimizers as it considers history of gradients, which allows to converge faster than other optimizers.
4. Momentum gradient descent is next to basic optimizer. It also considers history of gradients but is slower than Adam as it only calculates the moving average of the gradient and does not square it. This optimizer can be for any type of data set, as it is beginner level.