## **Running and Visualizing RMSProp**

Can we overcome aggressively decaying denominators?

1. Intuition: Why not decay the denominator and prevent its rapid growth?
2. We can consider the RMSProp algorithm
   * 1. Here we are taking an exponentially decaying sum
     2. Let and consider the 4th iteration
     3. We can see from this that our value is much smaller than in the case of Adagrad, due the history of the gradients being multiplied by the decay ratio.
     4. The relative difference between dense and sparse features is still maintained.
     5. This is the same as in Adagrad
3. Let’s visualise RMSProp in 2D
4. Adagrad got stuck when it was close to convergence (it was no longer able to move in the vertical (b) direction because of the decayed learning rate)
5. RMSProp overcomes this problem by being less aggressive on the decay