4712099 Optimization lecture Question

1. In Stochastic Gradient descent, the tuning parameter Ɛ (learning rate) is multiplied with the gradient g of minibatch of m examples and then subtracted from the learning parameter Θ which is updated iteratively till we reach a Ɛ_k.

$$\Theta \leftarrow \Theta - \varepsilon * g$$

What do you think the value of Ɛ should be, to arrive at a good instantiation?

   a. Value of Ɛ should be high always
   b. Value of Ɛ should be very less always
   c. The value of Ɛ should reduce / decay over time to arrive at a good instantiation
   d. Anyone of the above

Answer: c. The value of Ɛ should reduce / decay over time to arrive at a good instantiation