1 Learning rate decay

1.1 Definition

- slowly reduce the learning rate over time

1.2 Why

- afford taking larger steps during the initial steps of learning
- take smaller steps as learning approaches convergence

1.3 Formulation

$$\alpha = \frac{1}{1+decayrate * epochnum} * \alpha_0$$

1.4 Hyperparameters

- $decayrate$
- α_0

1.5 Variations

- exponential decay

  $$\alpha = 0.95^{epochnum} * \alpha_0$$

- constant decay

  $$\alpha = \frac{k}{\sqrt{epochnum}} * \alpha_0$$

  $$\alpha = \frac{k}{\sqrt{t}} * \alpha_0$$

- discrete staircase decay