PadhAI: Variants of Gradient Descent

One Fourth Labs

A limitation of Adagrad

What do we observe?

1. **Advantage**: Parameters corresponding to sparse features get better updates
2. **Disadvantage**: The learning rate decays very aggressively as the denominator grows (not good for parameters corresponding to dense features)