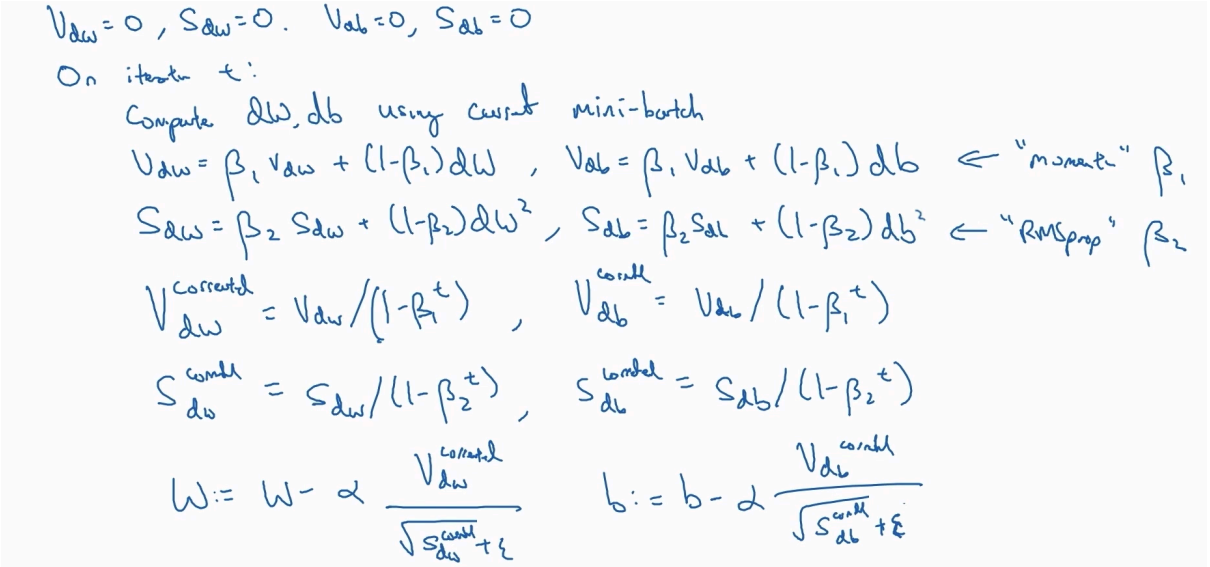
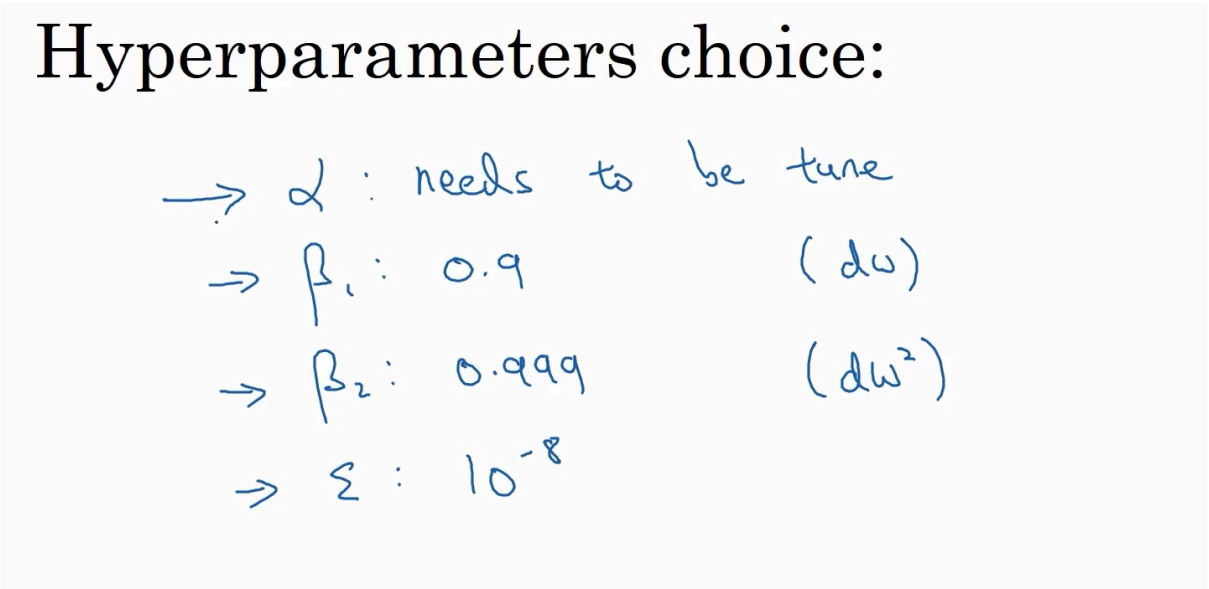
Adam Optimizer

Adam = RMSProp + Momentum (Adaptive Moment Estimation)

- Start by initializing V_dW, S_dW and V_db, S_db to all zero
- Then do the Momentum calculations for V_dW, V_db
- Then do the RMS calculations for S_dW, S_db
- Then do bias corrections for all
- Then just combine Momentum update with RMS update 😀





Only need to tune alpha usually. b1 and b2 were the default suggestions by the authors of the Adam optimizer paper