

Optimization Algorithms

- ✓ **Video:** Mini-batch Gradient Descent
11 min
- ✓ **Video:** Understanding Mini-batch Gradient Descent
11 min
- ✓ **Video:** Exponentially Weighted Averages
5 min
- ✓ **Video:** Understanding Exponentially Weighted Averages
9 min
- ✓ **Video:** Bias Correction in Exponentially Weighted Averages
4 min
- ✓ **Video:** Gradient Descent with Momentum
9 min
- ✓ **Video:** RMSprop
7 min
- 📖 **Reading:** Clarification about Upcoming Adam Optimization Video
1 min
- ▶ **Video:** Adam Optimization Algorithm
7 min
- 📖 **Reading:** Clarification about Learning Rate Decay Video
1 min
- ▶ **Video:** Learning Rate Decay
6 min
- ▶ **Video:** The Problem of Local Optima
5 min

Lecture Notes (Optional)

Quiz

Programming Assignment

Heroes of Deep Learning (Optional)

Clarification about Upcoming Adam Optimization Video

Please note that in the next video at 2:44, the following picture is correct. However, later in the video, the db^2 lost the 2 .

The bottom right equation should be:

$$S_{db} = \beta_2 S_{db} + (1 - \beta_2) db^2$$

Adam optimization algorithm

$V_{dw} = 0, S_{dw} = 0, V_{db} = 0, S_{db} = 0$
On iteration t :
Compute dw, db using current mini-batch
 $V_{dw} = \beta_1 V_{dw} + (1 - \beta_1) dw, V_{db} = \beta_1 V_{db} + (1 - \beta_1) db \leftarrow$
 $S_{dw} = \beta_2 S_{dw} + (1 - \beta_2) dw^2, S_{db} = \beta_2 S_{db} + (1 - \beta_2) db^2 \leftarrow$

▶ 2:44 / 7:07

⌵ 🔊 ⚙️ andrew Ng

Mark as completed

