Optimization

Outline

- Recap minibatch SGD
- Discuss semester timeline
- Some prerequisites (SGD, contour plots, EMA)
- Momentum
- Adaptive learning rate
  - AdaGrad (Adaptive Gradients)
  - RMSProp (Root-Mean-Square Backpropagation)
  - Adam (Adaptive Moment Estimation)

Recap: Minibatch SGD

- Take five minutes to draw
  - Whatever will help you remember (no correct or incorrect drawings)
  - You'll keep a running drawing log the rest of the semester

Semester vs. Topic Timelines

Foundations

- Math
  - Calculus
  - Linear algebra
- Programming
  - Python
  - Notebooks
  - Libraries
- Computing
  - HPC
  - CLI

NN Basics

- Terminology
- History
- Ethics
- Neurons
- Networks
- Backpropagation
- Autodiff
- Gradient descent

NN Intermediate

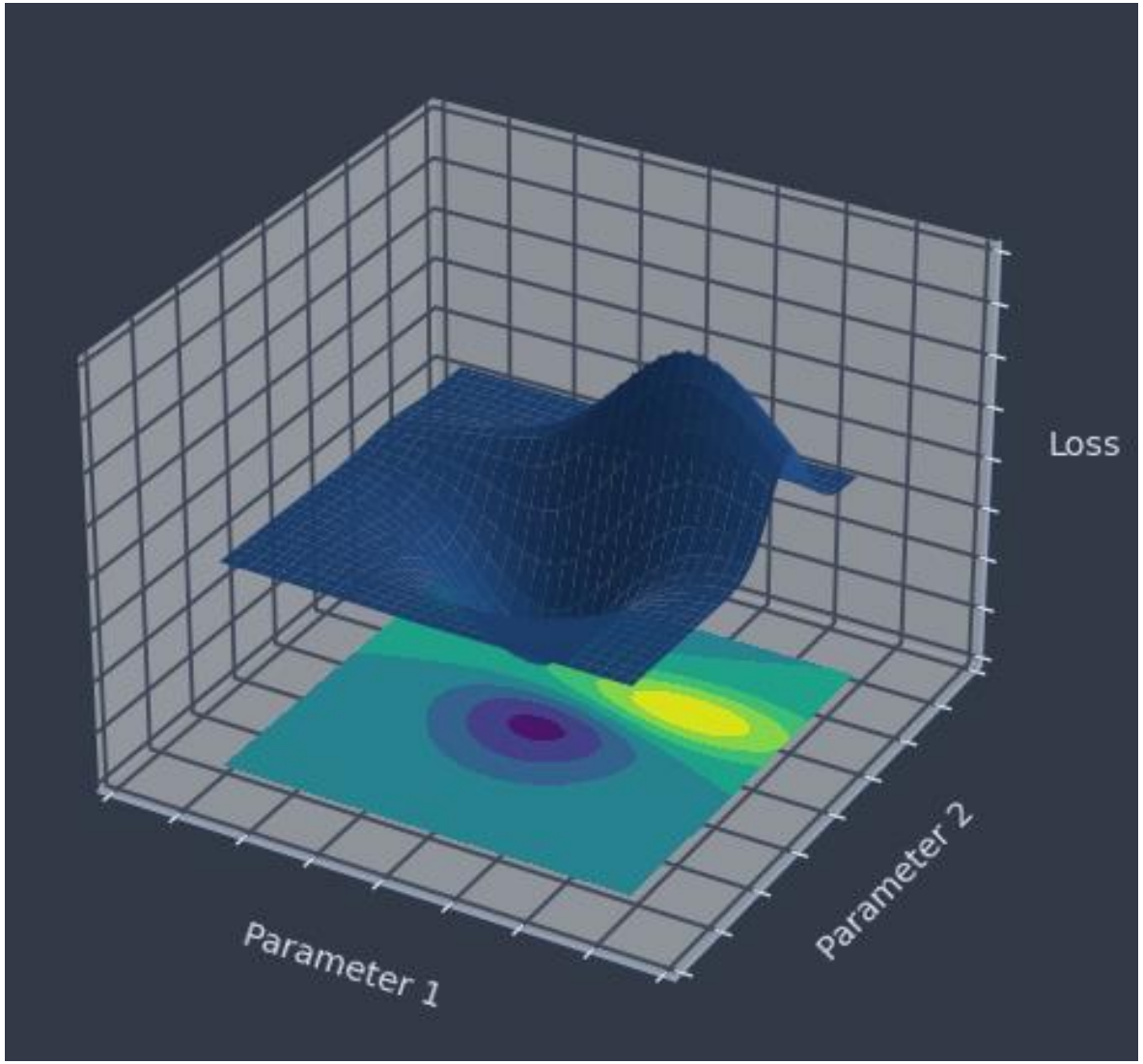
- Optimization
- Initialization
- Normalization
- Overfitting
- Convolutions
- Recurrent

NN Advanced

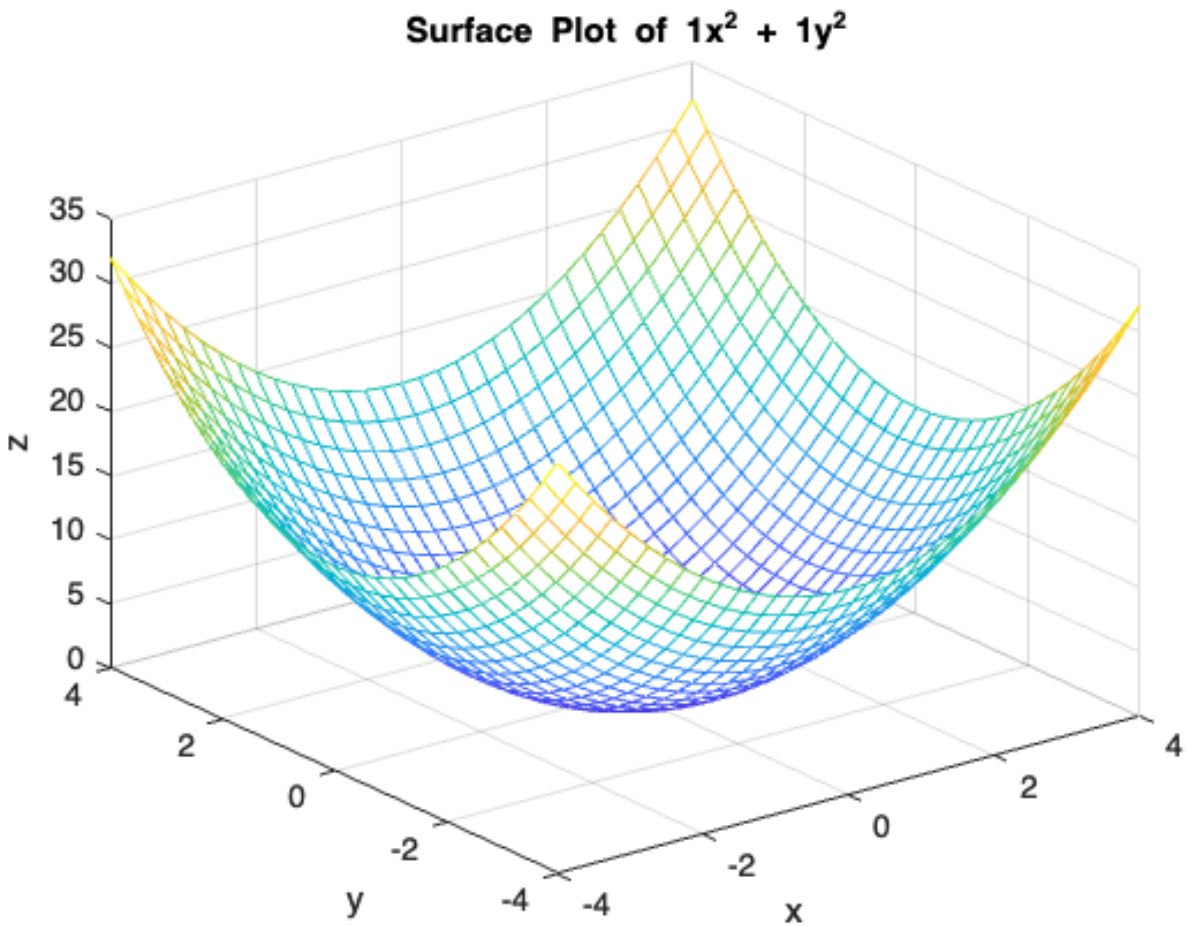
- Transfer learning
- Inference
- GANs
- Reinforcement
- Attention
- Transformers
- Stable diffusion

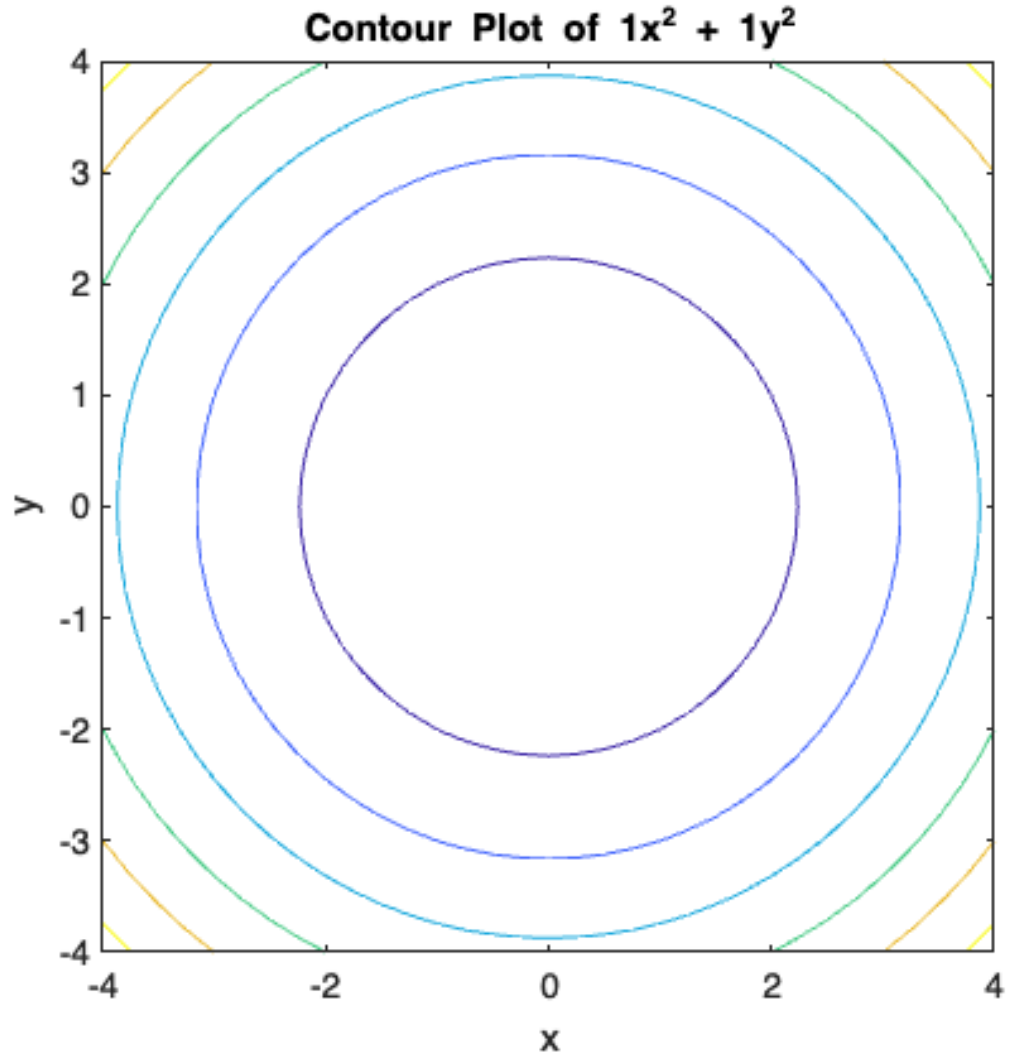
SGD Equations

Contour Plots

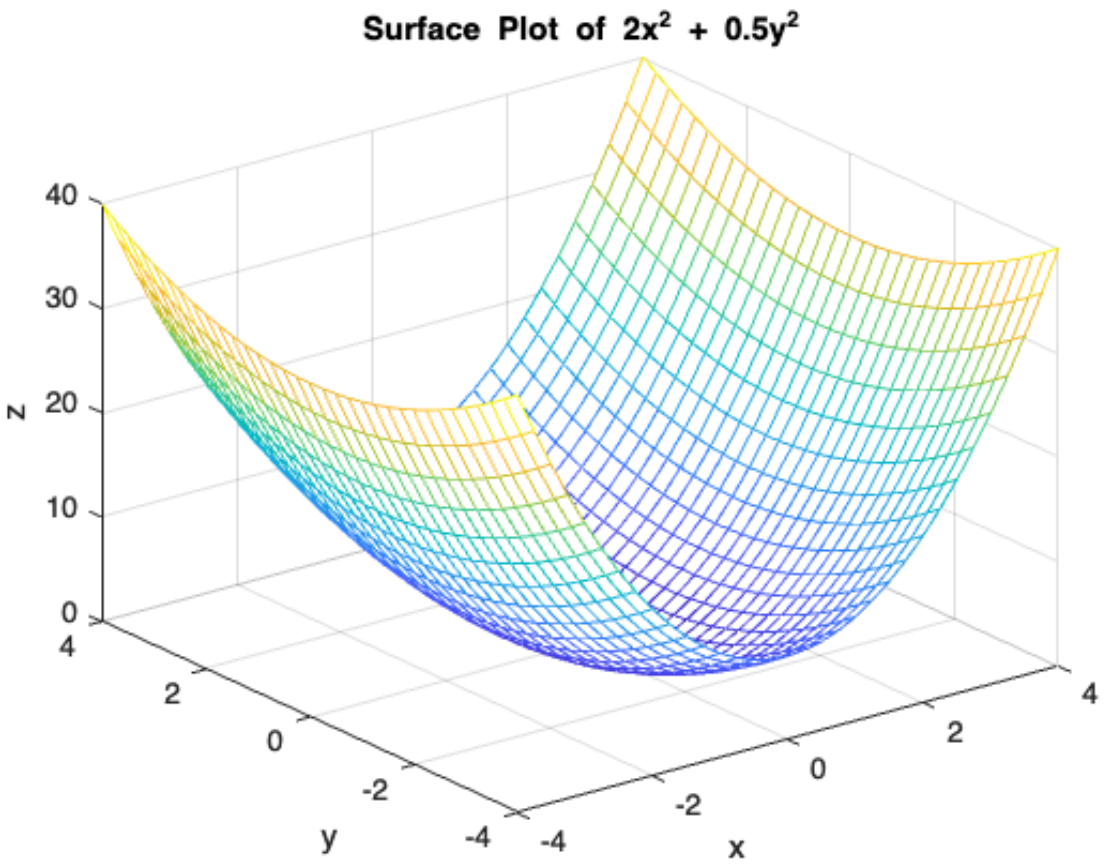


Uniform Gradients

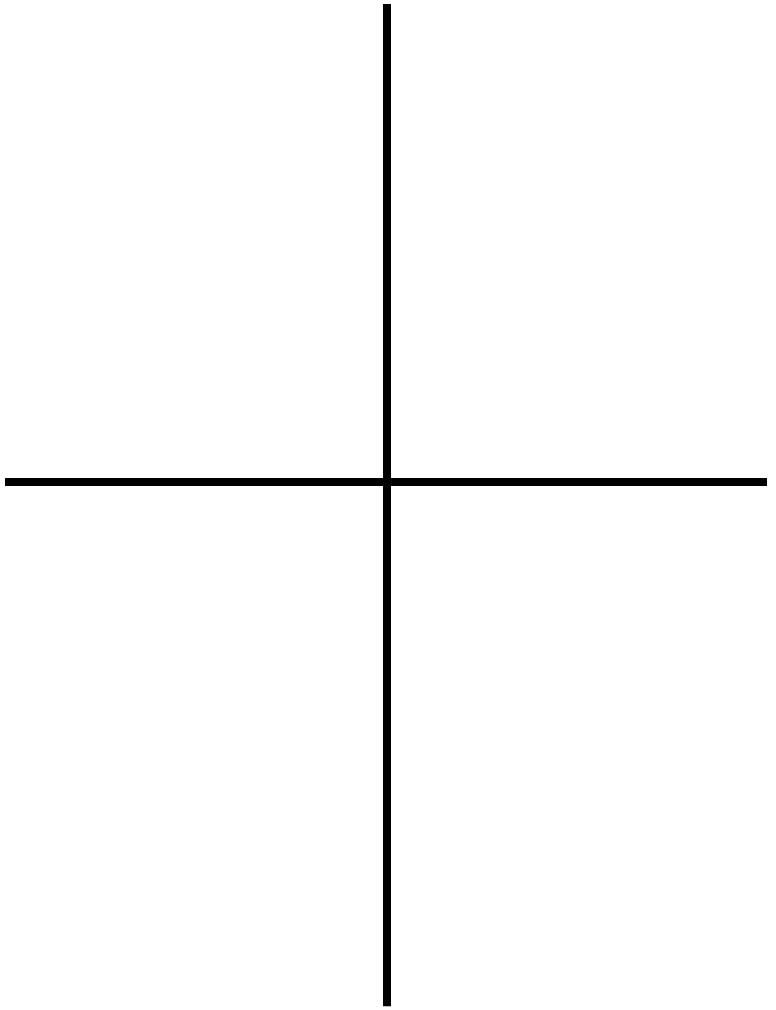




Contour Plots



Contour Plot



Exponential Moving Average (EMA)

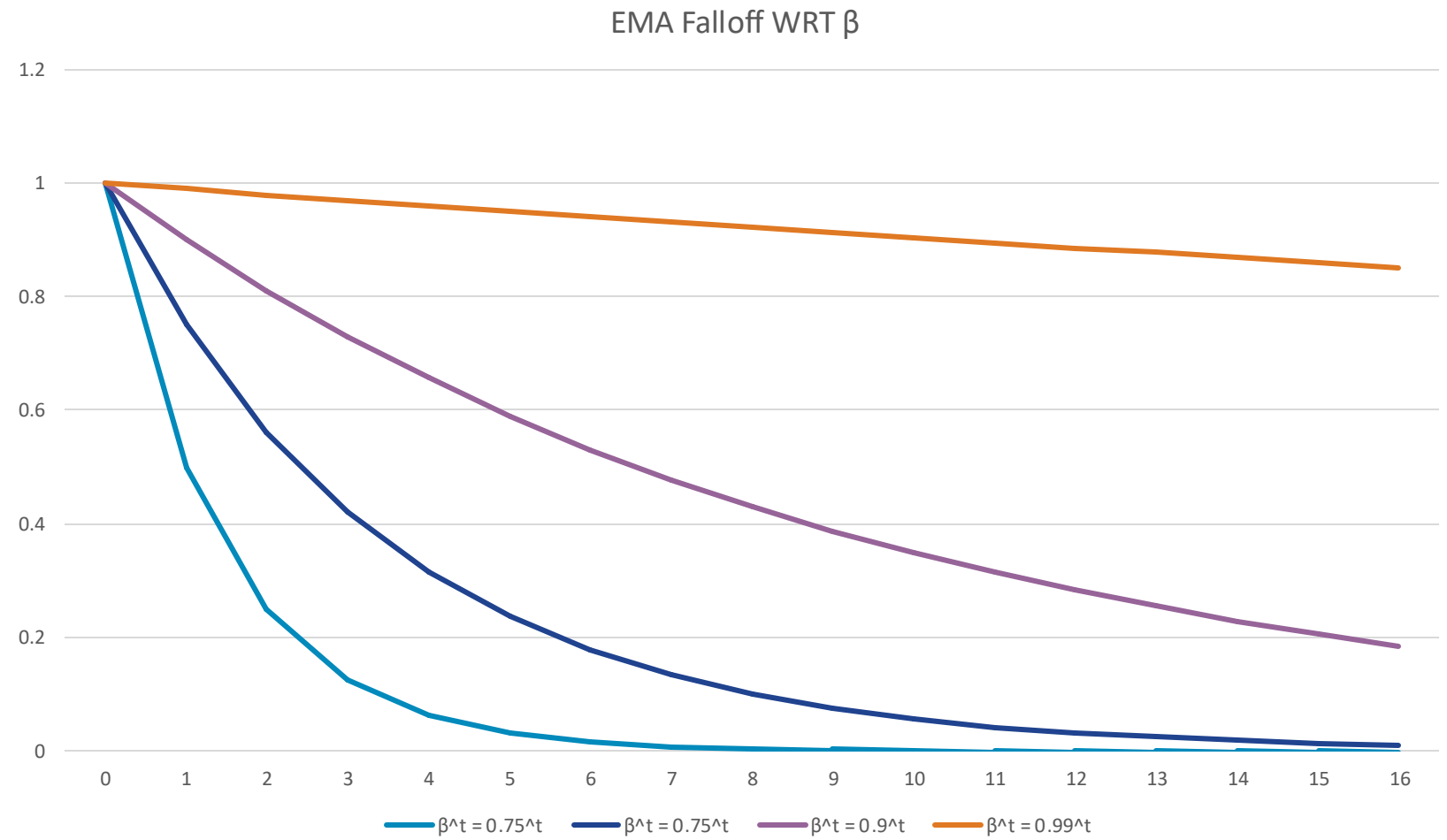
- A simple moving average is the unweighted mean of the previous k values
- EMA is a rule-of-thumb technique for smoothing time series data

Exponential Moving Average (EMA)

- https://cs.pomona.edu/classes/cs181r/book/19-SensorFusion.html

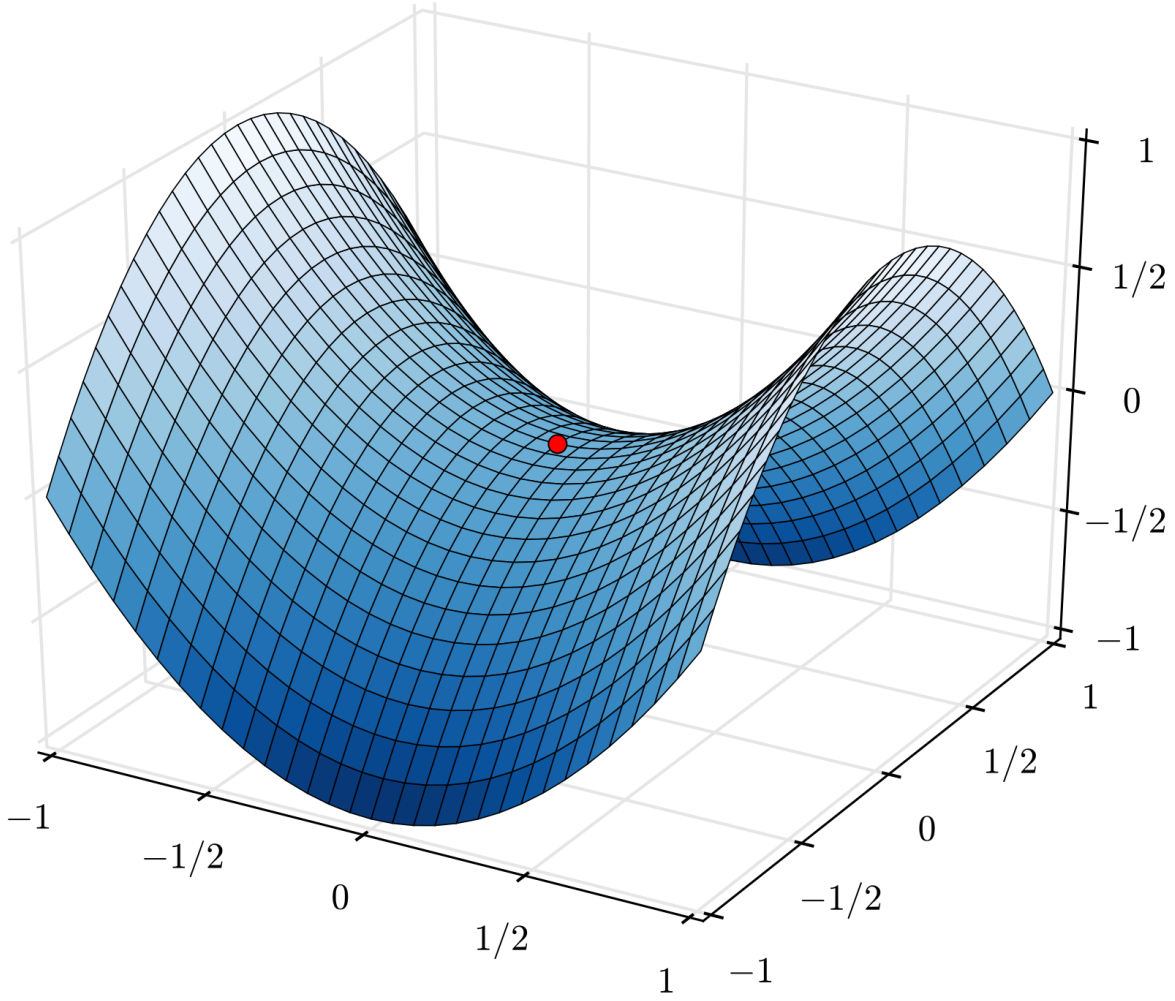
Exponential Moving Average (EMA)

- A simple moving average is the unweighted mean of the previous k values
- EMA is a rule-of-thumb technique for smoothing time series data

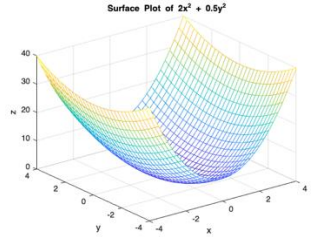


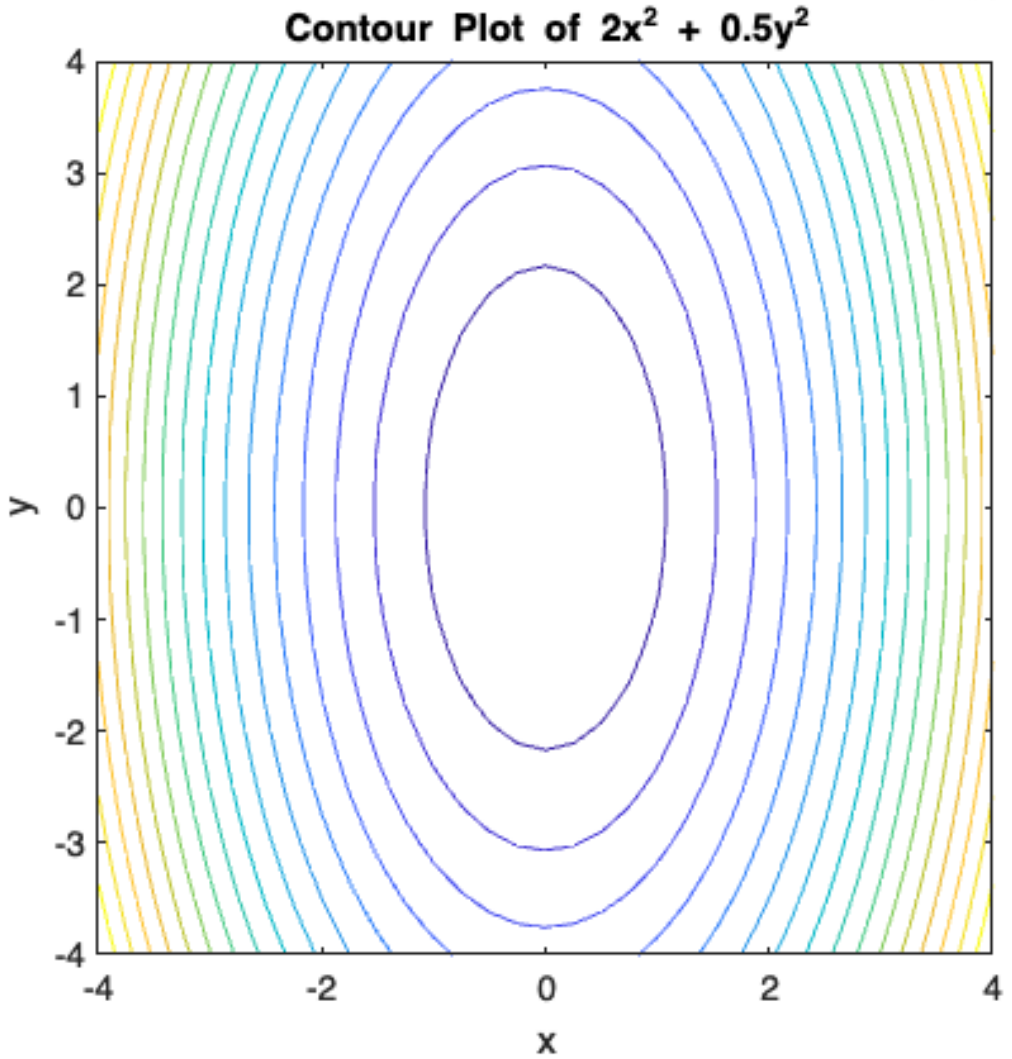
Gradient Plateau

Saddle Points

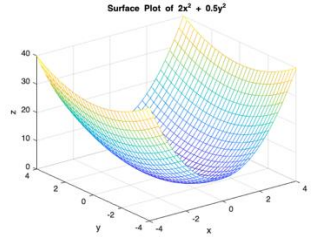


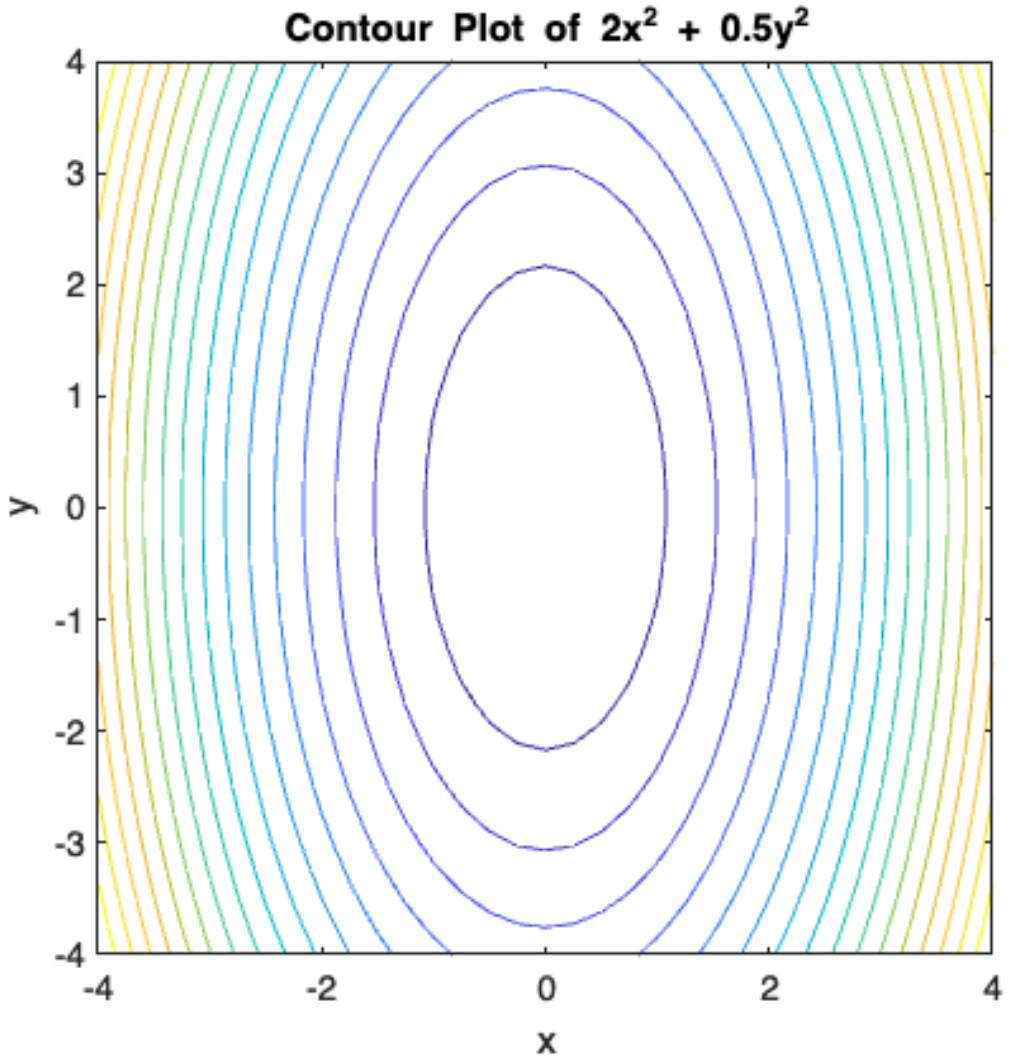
Momentum


Surface Plot of 2x² + 0.5y²

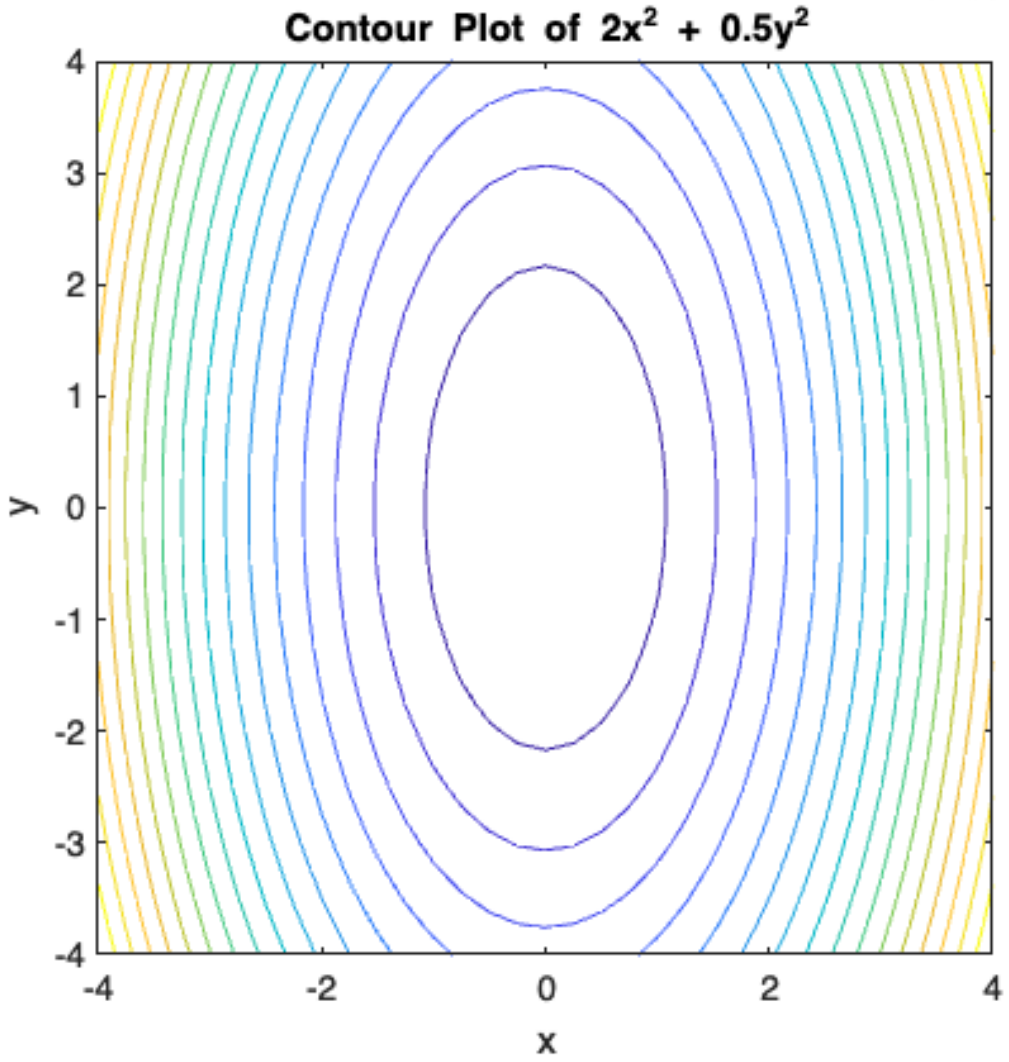

Contour Plot of 2x² + 0.5y²
y
x
4
3
2
1
0
-1
-2
-3
-4
-4
-2
0
2
4

Momentum





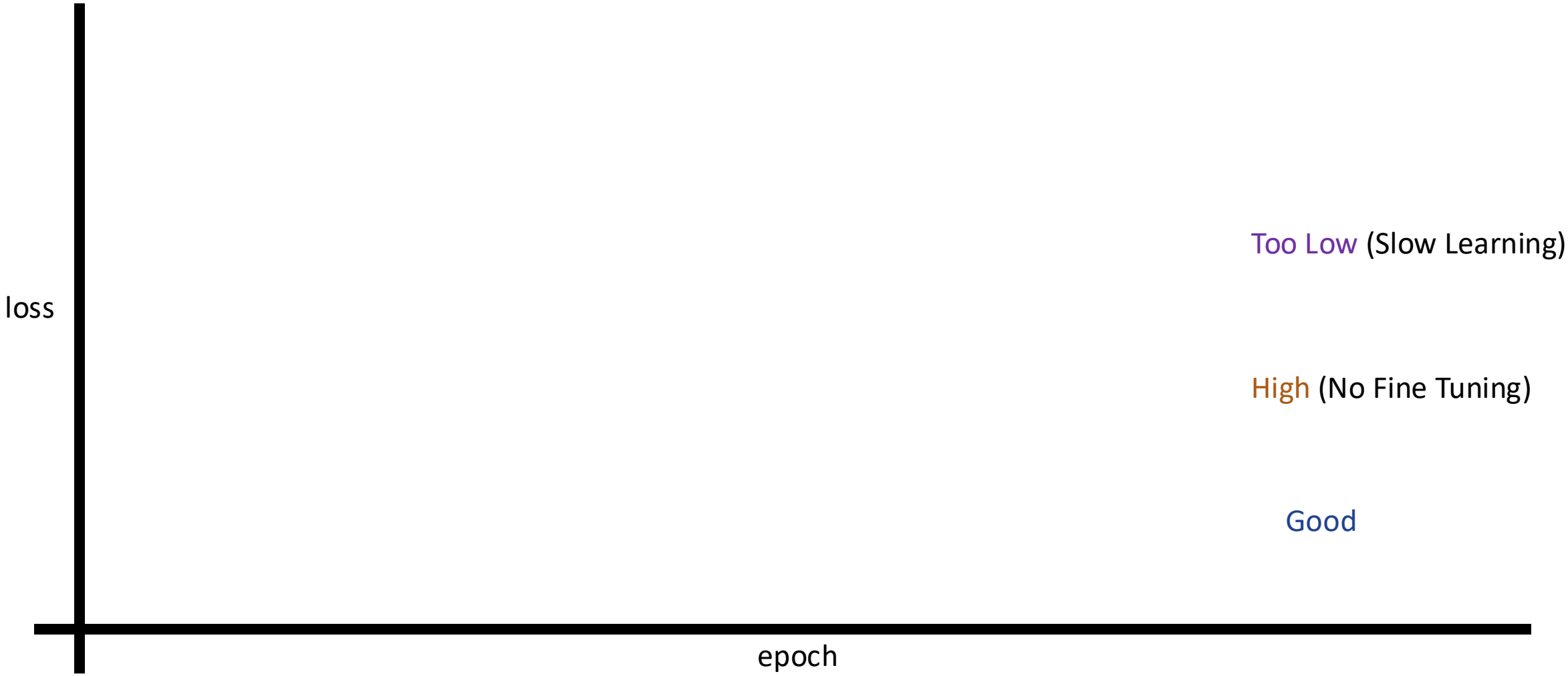
Without Momentum



With Momentum

Learning Rates

Too High (No Learning)



Too Low (Slow Learning)

High (No Fine Tuning)

Good

AdaGrad

RMSProp

Adam

Recent Optimizers

- SGD (1951)
- SGD+Momentum (1999)
- AdaGrad (2011)
- AdaDelta (2012)
- RMSProp (2013)
- Adam (2014)
- NADAM (2015)
- AdamW (2017)
- AdaShift (2018)
- AggMo (2018)
- LAMB (2019)
- AMSGrad (2019)
- Adabelief (2020)
- MADGRAD (2021)
- AdaSmooth (2022)

Just a sample

See: https://johnchenresearch.github.io/demon/ for more information