Gradient Descent with Momentum
RMSprop
Adam

4th December 2023

Overview

Problem with gradient descent

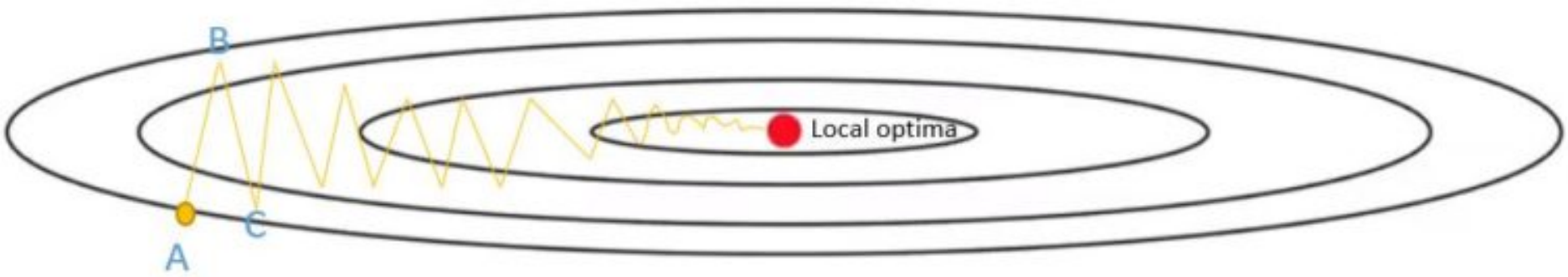


Figure: Gradient descent [1]

[1] https://engmrk.com/gradient-descent-with-momentum/.

Problem with gradient descent

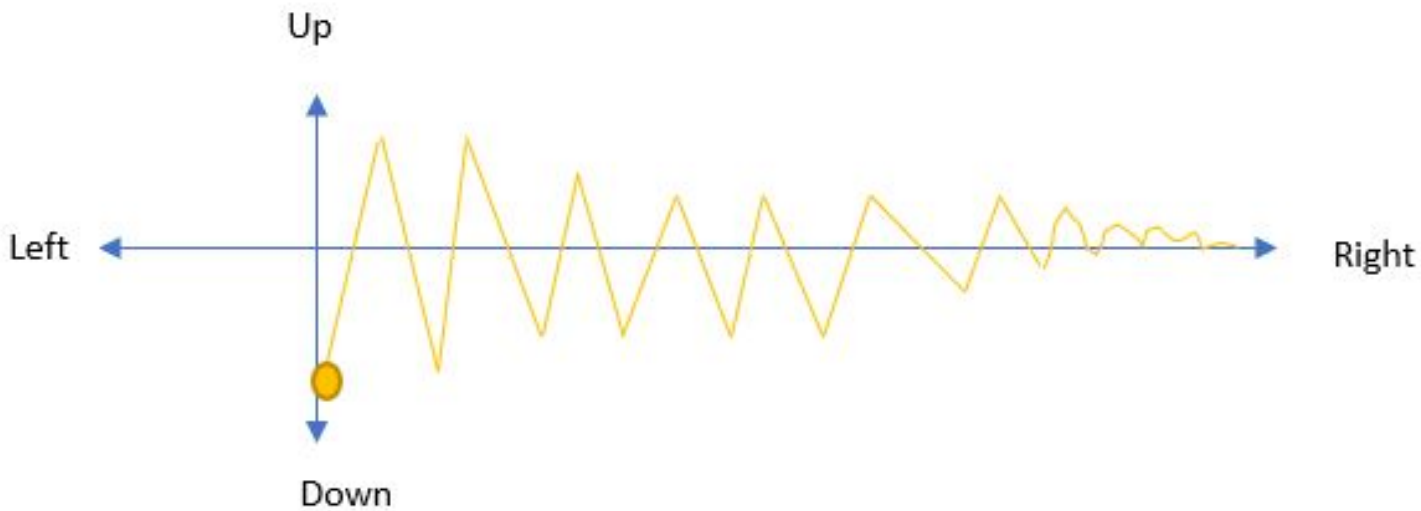
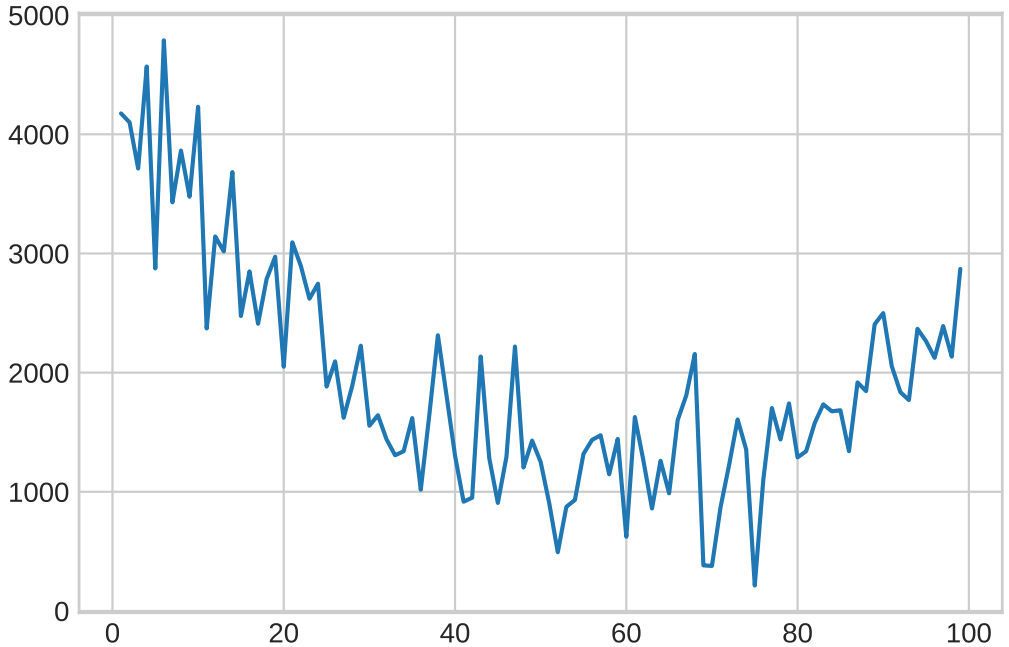


Figure: Gradient descent [2]

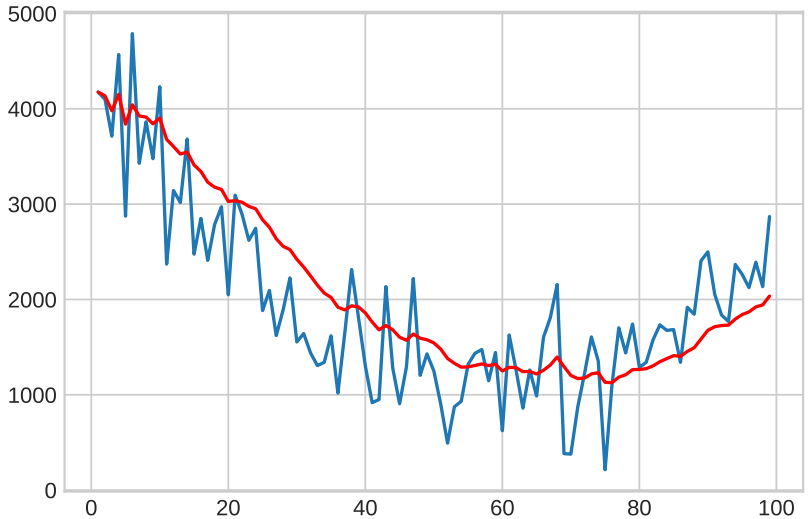
[2] https://engmrk.com/gradient-descent-with-momentum/.

Exponentially weighted moving average

Example: How to model/forecast the stock/share price of a certain company?

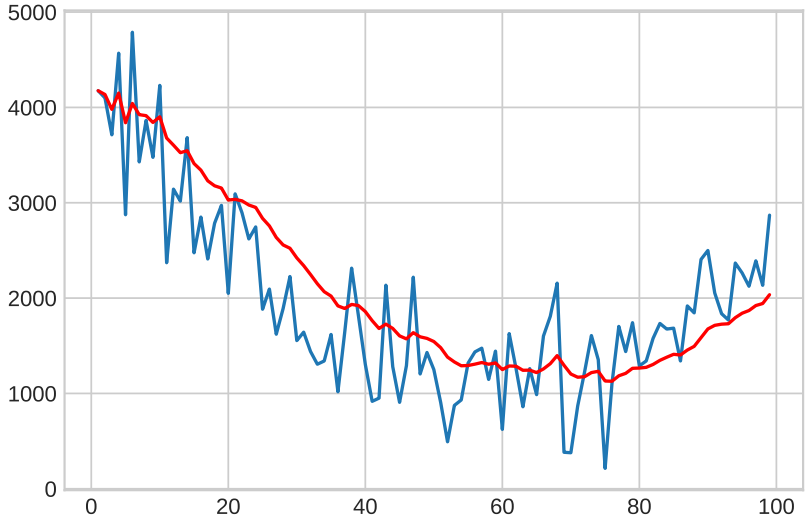


Exponentially weighted moving average



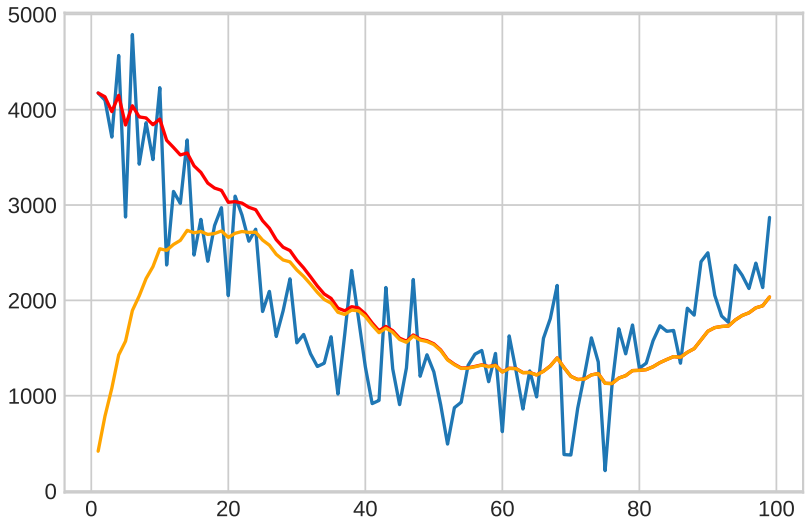
$$m_{100} = \beta m_{99} + (1 - \beta) price_{100}$$
$$m_{99} = \beta m_{98} + (1 - \beta) price_{99}$$
$$\dots$$
$$m_t = \beta m_{t-1} + (1 - \beta) price_t$$
$$\dots$$
$$m_1 = \beta m_0 + (1 - \beta) price_1$$

Exponentially weighted moving average



$$
\begin{aligned}
\beta &= 0.9 \\
m_{100} &= 0.9 m_{99} + (1 - 0.9) price_{100} \\
&= 0.1 price_{100} + 0.9 m_{99} \\
&= 0.1 price_{100} + 0.9(0.1 price_{99} + 0.9 m_{98}) \\
&= 0.1 price_{100} + 0.9 \cdot 0.1 price_{99} + (0.9)^2 m_{98} \\
&= 0.1 price_{100} + 0.09 price_{99} + (0.9)^2(0.1 price_{98} + 0.9 m_{97}) \\
&= 0.1 price_{100} + 0.09 price_{99} + 0.081 price_{98} + (0.9)^3 m_{97} \\
&= \ldots
\end{aligned}
$$

Exponentially weighted moving average



Actually, without the below correction, we will get the orange line instead of the red line. Why?

$$m_t = \beta m_{t-1} + (1-\beta) price_t$$

$$m_t^{corrected} = \frac{m_t}{1-\beta^t}$$

Gradient descent with momentum

Update weights at iteration t:

$$\begin{aligned} \mathrm{g}_{\{t\}} &\longleftarrow \nabla L(\mathrm{W}_{\{t-1\}}) \\ \mathrm{m}_{\{t\}} &\longleftarrow \beta_1 \mathrm{m}_{\{t-1\}} + (1-\beta_1)\mathrm{g}_{\{t\}} \\ \mathrm{W}_{\{t\}} &\longleftarrow \mathrm{W}_{\{t-1\}} - \gamma \mathrm{m}_{\{t\}} \end{aligned}$$

Initialization: $\mathrm{m}_{\{0\}} = 0$

Hyperparameters: $\beta_1 = 0.9$; learning rate γ

Gradient descent with momentum

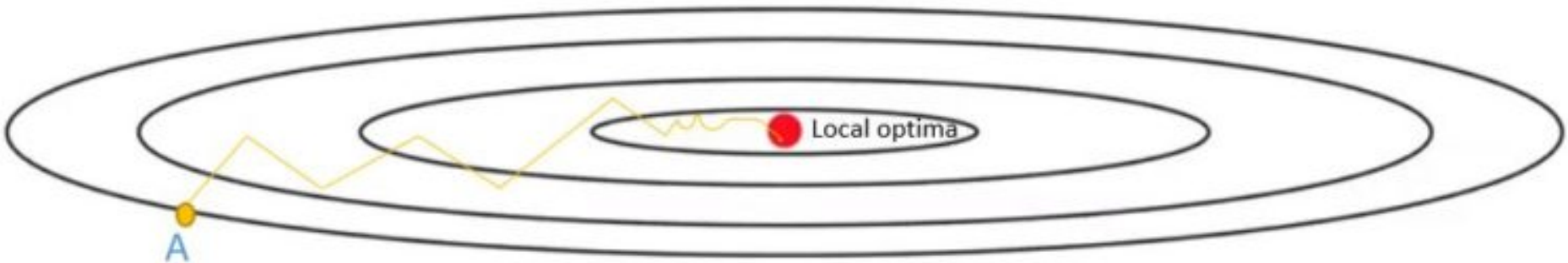


Figure: Gradient descent with momentum [3]

[3] https://engmrk.com/gradient-descent-with-momentum/.

Problem with gradient descent

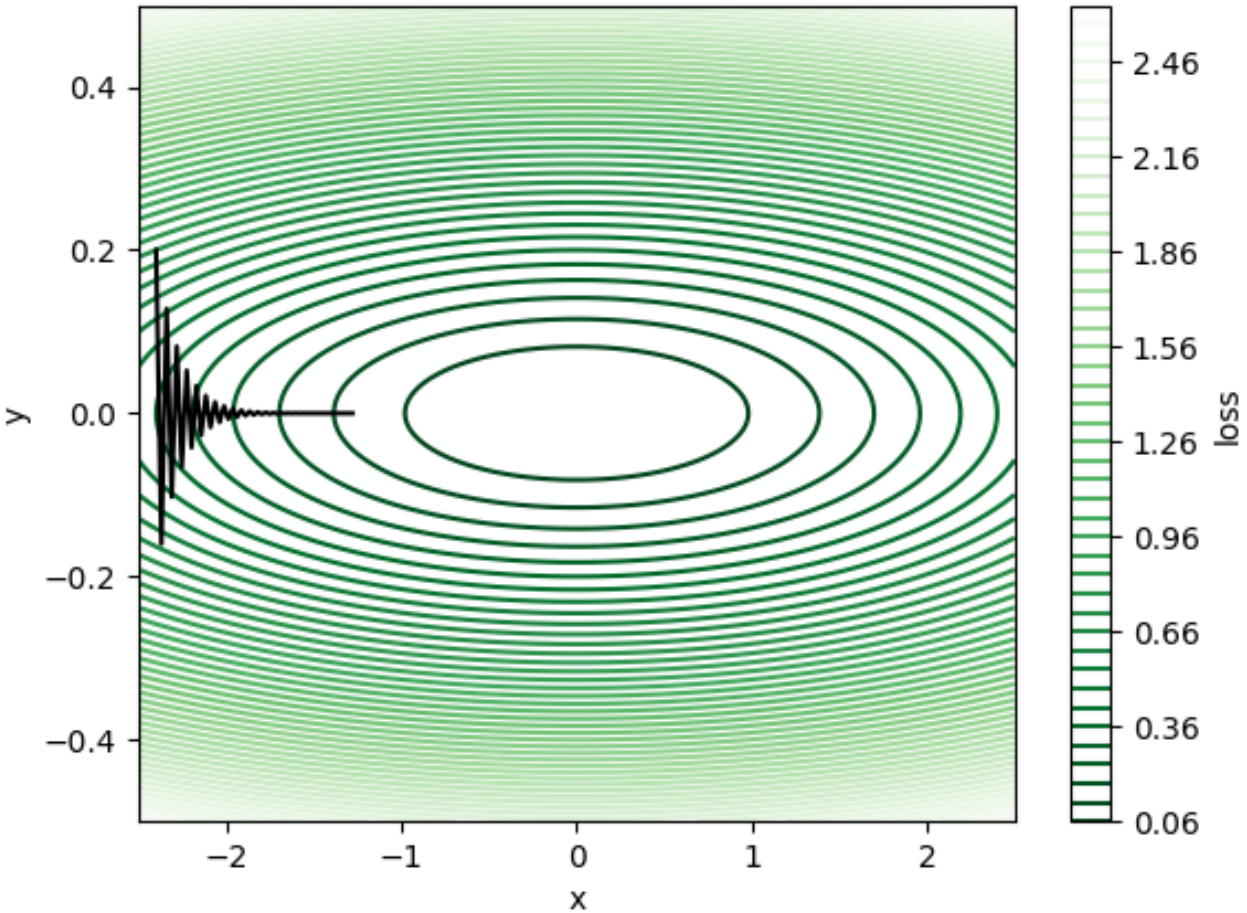


Figure: Gradient descent [4]

[4] https://github.com/hengluchang/visualizing_momentum.

Gradient descent with momentum

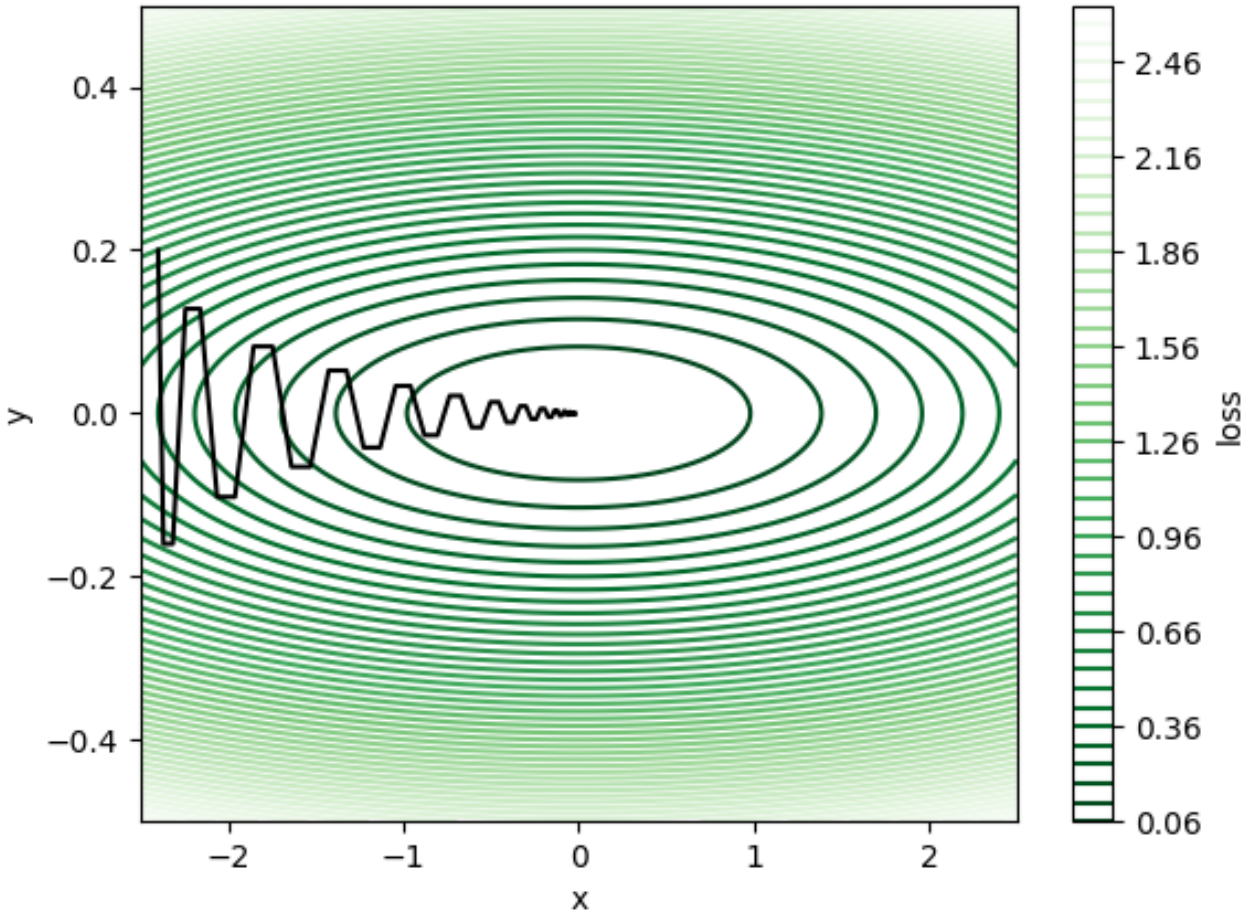


Figure: Gradient descent with momentum $\beta_1 = 0.8$ [5]

[5] `https://github.com/hengluchang/visualizing_momentum`.

RMSprop

Update weights at iteration t:

$$\begin{aligned} \mathsf{g}_{\{t\}} &\longleftarrow \nabla L(\mathsf{W}_{\{t-1\}}) \\ \mathsf{s}_{\{t\}} &\longleftarrow \beta_2 \mathsf{s}_{\{t-1\}} + (1-\beta_2)\mathsf{g}_t^2 \\ \mathsf{W}_{\{t\}} &\longleftarrow \mathsf{W}_{\{t-1\}} - \gamma \frac{\mathsf{g}_{\{t\}}}{\sqrt{\mathsf{s}_{\{t\}}} + \epsilon} \end{aligned}$$

Initialization: $\mathsf{s}_{\{0\}} = 0$

Hyperparameters: $\beta_2 = 0.9$; $\epsilon = 10^{-8}$; learning rate γ

Adam - Adaptive moment estimation

Update weights at iteration t:

$$
\begin{aligned}
\mathrm{g}_{\{t\}} &\longleftarrow \nabla L(\mathrm{W}_{\{t-1\}}) \\
\mathrm{m}_{\{t\}} &\longleftarrow \beta_1 \mathrm{m}_{\{t-1\}} + (1-\beta_1)\mathrm{g}_{\{t\}} \\
\mathrm{s}_{\{t\}} &\longleftarrow \beta_2 \mathrm{s}_{\{t-1\}} + (1-\beta_2)\mathrm{g}_t^2 \\
\mathrm{m}_{\{t\}}^{corrected} &\longleftarrow \frac{\mathrm{m}_{\{t\}}}{1-\beta_1^t} \\
\mathrm{s}_{\{t\}}^{corrected} &\longleftarrow \frac{\mathrm{s}_{\{t\}}}{1-\beta_2^t} \\
\mathrm{W}_{\{t\}} &\longleftarrow \mathrm{W}_{\{t-1\}} - \gamma \frac{\mathrm{m}_{\{t\}}^{corrected}}{\sqrt{\mathrm{s}_{\{t\}}^{corrected}} + \epsilon}
\end{aligned}
$$

Initialization: $\mathrm{m}_{\{0\}} = \mathrm{s}_{\{0\}} = 0$

Hyperparameters: $\beta_1 = 0.9$; $\beta_2 = 0.999$; $\epsilon = 10^{-8}$; learning rate γ

References

1. Visualizing Gradient Descent with Momentum in Python: https://github.com/hengluchang/visualizing_momentum.
2. Coursera - Improving Deep Neural Networks: Hyperparameter tuning, Regularization and Optimization: https://www.coursera.org/learn/deep-neural-network/
3. Coursera - Neural Networks and Deep Learning: https://www.coursera.org/learn/neural-networks-deep-learning/