# Deep Learning

Vazgen Mikayelyan

YSU, Krisp

October 28, 2020

# Outline

# Gradient Descent with Momentum

Let $L(w)$ be a loss function that we want to minimize. The algorithm gradient descent with momentum is the following

$$v_0 = 0,$$

$$v_t = \beta v_{t-1} + (1-\beta)\nabla L(w_t),$$

$$w_{t+1} = w_t - \alpha v_t,$$

where $\alpha$ is the learning rate and $\beta \in [0, 1)$ is the parameter of exponential moving average.

# Gradient Descent with Momentum

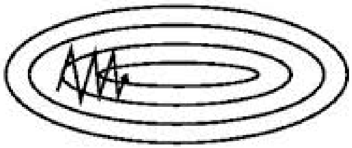

Image 2: SGD without momentum

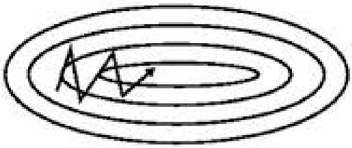

Image 3: SGD with momentum

# RMSProp

Let $L(w)$ be a loss function that we want to minimize. The algorithm RMSprop is the following

$$v_0 = 0,$$
$$v_t = \beta v_{t-1} + (1-\beta)\left(\nabla L(w_t)\right)^2,$$
$$w_{t+1} = w_t - \alpha \frac{\nabla L(w_t)}{\sqrt{v_t} + \varepsilon},$$

where $\alpha$ is the learning rate and $\beta \in [0,1)$ is the parameter of exponential moving average.

# ADAM

Let $L(w)$ be a loss function that we want to minimize. The algorithm ADAM is the following

$$m_0 = 0, v_0 = 0,$$

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) \nabla L(w_{t-1}),$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t},$$

# ADAM

Let $L(w)$ be a loss function that we want to minimize. The algorithm ADAM is the following

$$m_0 = 0, v_0 = 0,$$

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) \nabla L(w_{t-1}),$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t},$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) \left(\nabla L(w_{t-1})\right)^2,$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t},$$

# ADAM

Let $L(w)$ be a loss function that we want to minimize. The algorithm ADAM is the following

$$m_0 = 0, v_0 = 0,$$

$$m_t = \beta_1 m_{t-1} + (1-\beta_1)\nabla L(w_{t-1}),$$

$$\hat{m}_t = \frac{m_t}{1-\beta_1^t},$$

$$v_t = \beta_2 v_{t-1} + (1-\beta_2)\left(\nabla L(w_{t-1})\right)^2,$$

$$\hat{v}_t = \frac{v_t}{1-\beta_2^t},$$

$$w_t = w_{t-1} - \alpha\frac{\hat{m}_t}{\sqrt{\hat{v}_t}+\varepsilon},$$

where $\alpha$ is the learning rate and $\beta_1, \beta_2 \in [0,1)$ are the parameters of exponential moving averages.