Problems with Gradient Descent

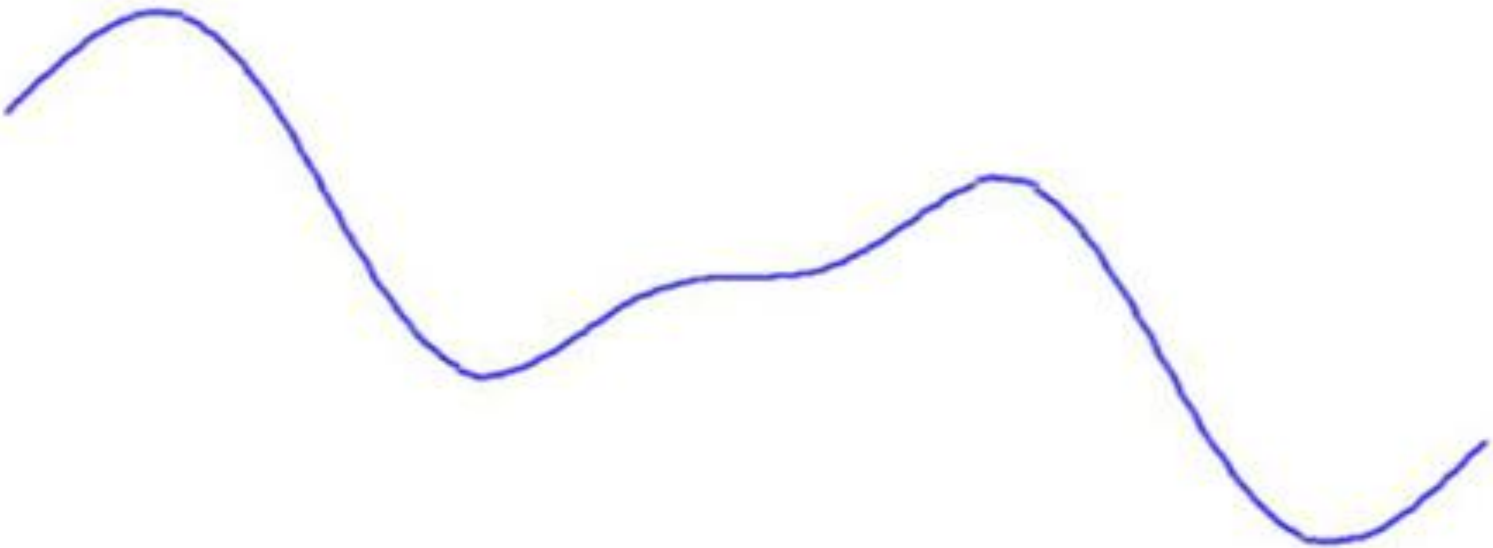
Problems with Gradient Descent

1. Getting stuck at local minima

Analytics
Vidhya

Problems with Gradient Descent

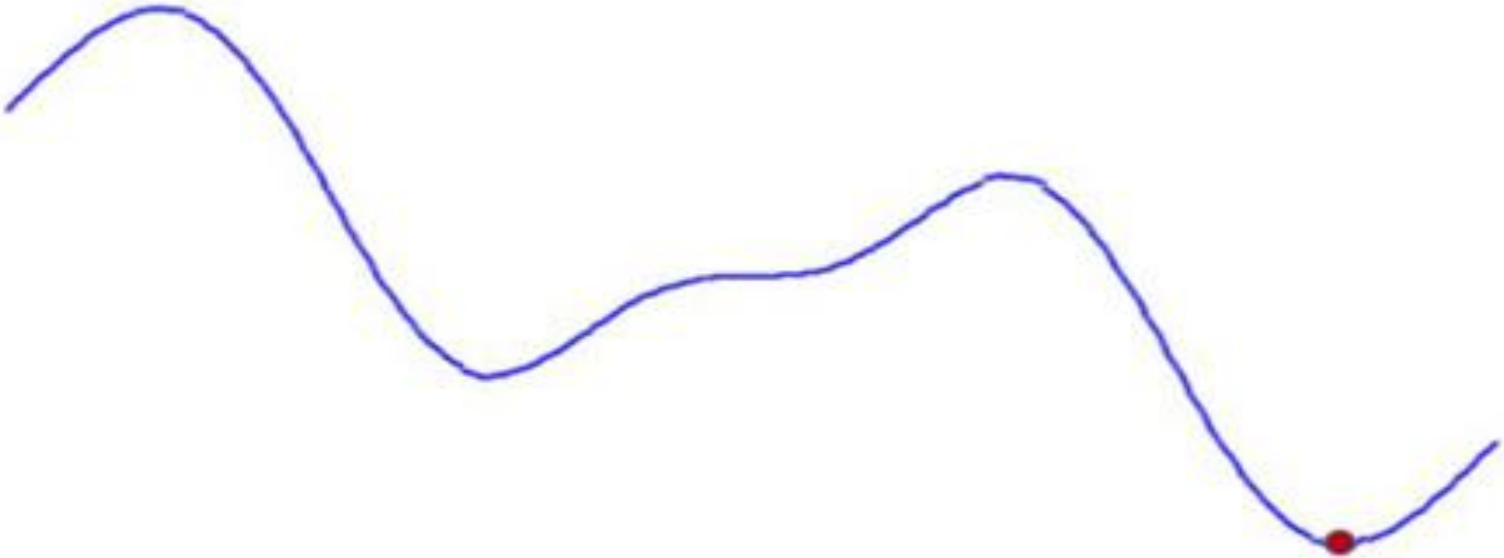
1. Getting stuck at local minima



Analytics Vidhya
Learn everything about analytics

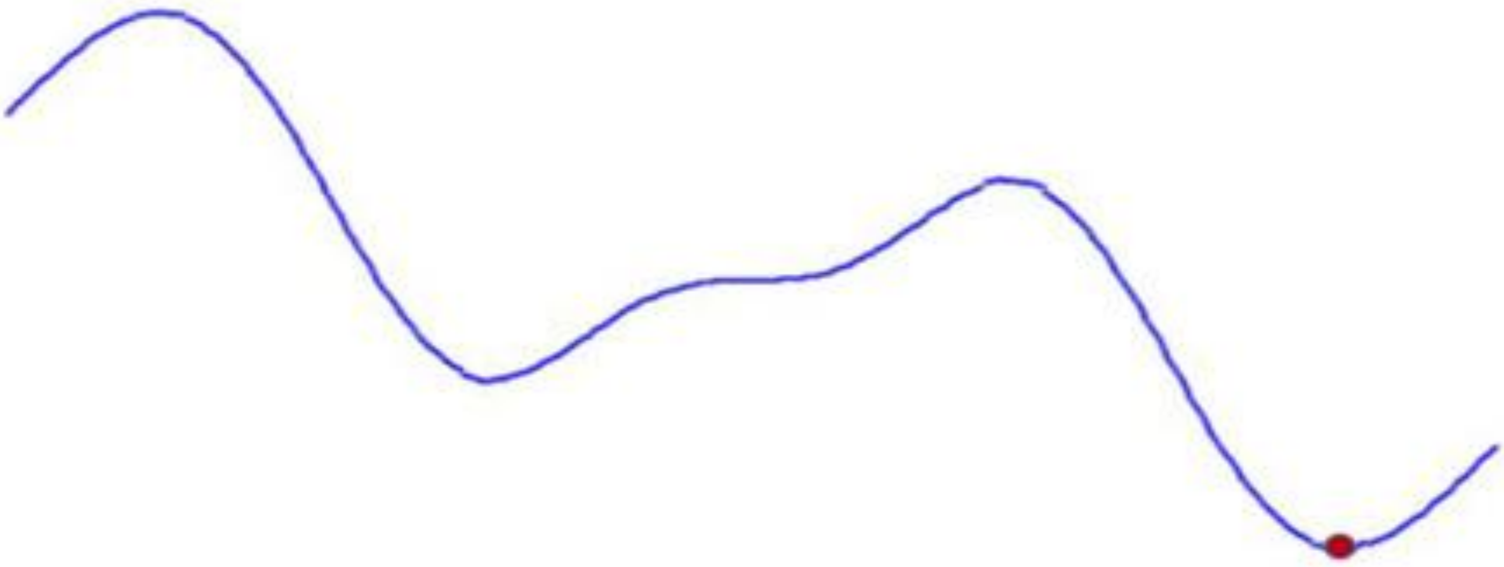
Problems with Gradient Descent

1. Getting stuck at local minima



Problems with Gradient Descent

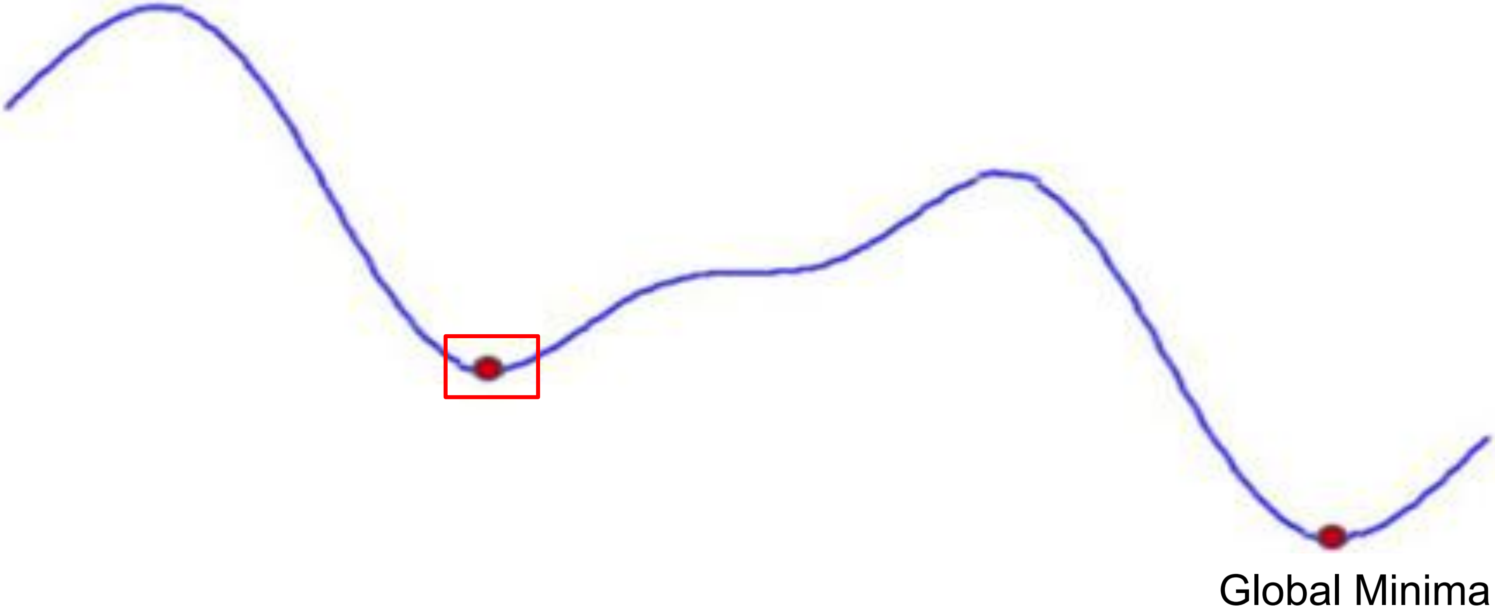
1. Getting stuck at local minima



Global Minima

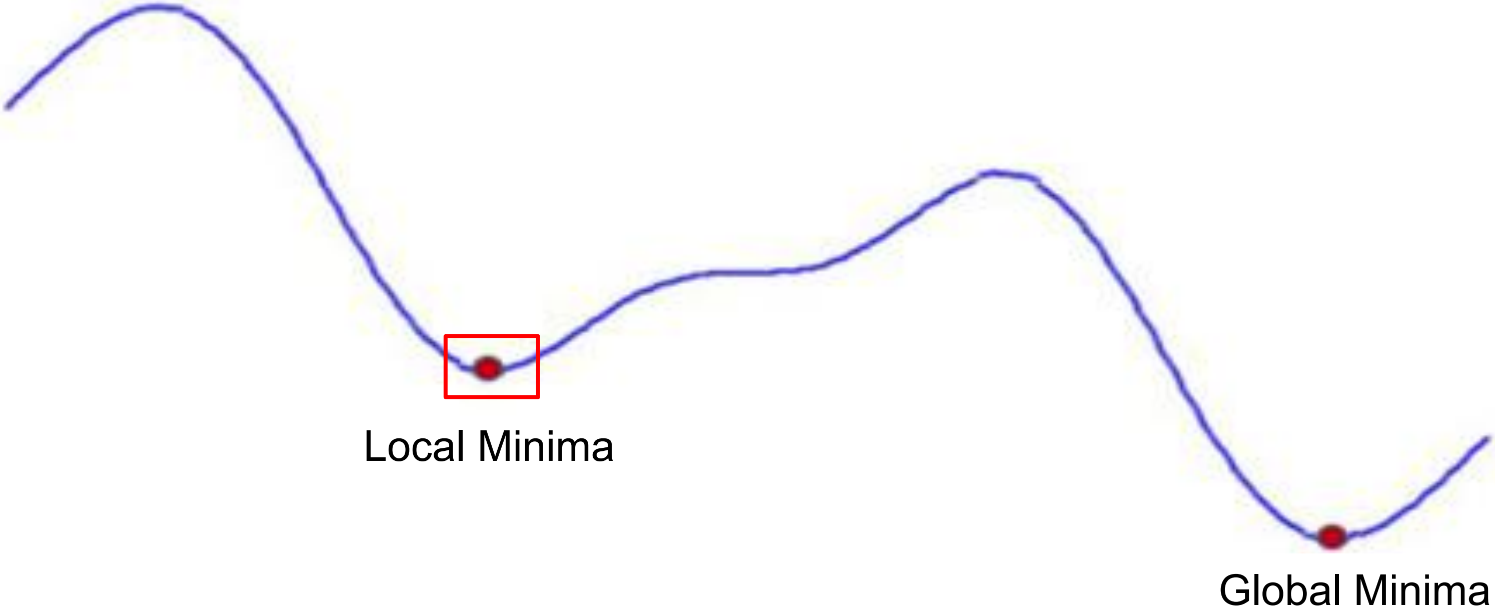
Problems with Gradient Descent

1. Getting stuck at local minima


Global Minima

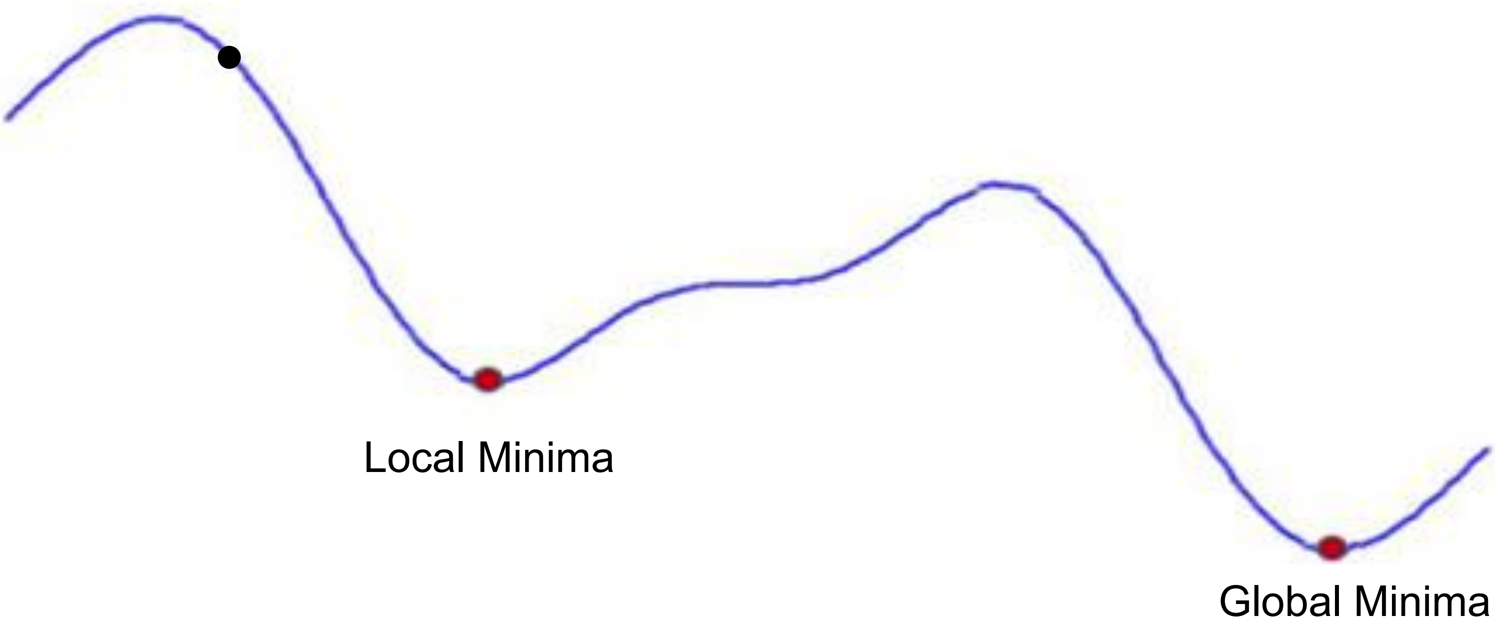
Problems with Gradient Descent

1. Getting stuck at local minima


Local Minima
Global Minima

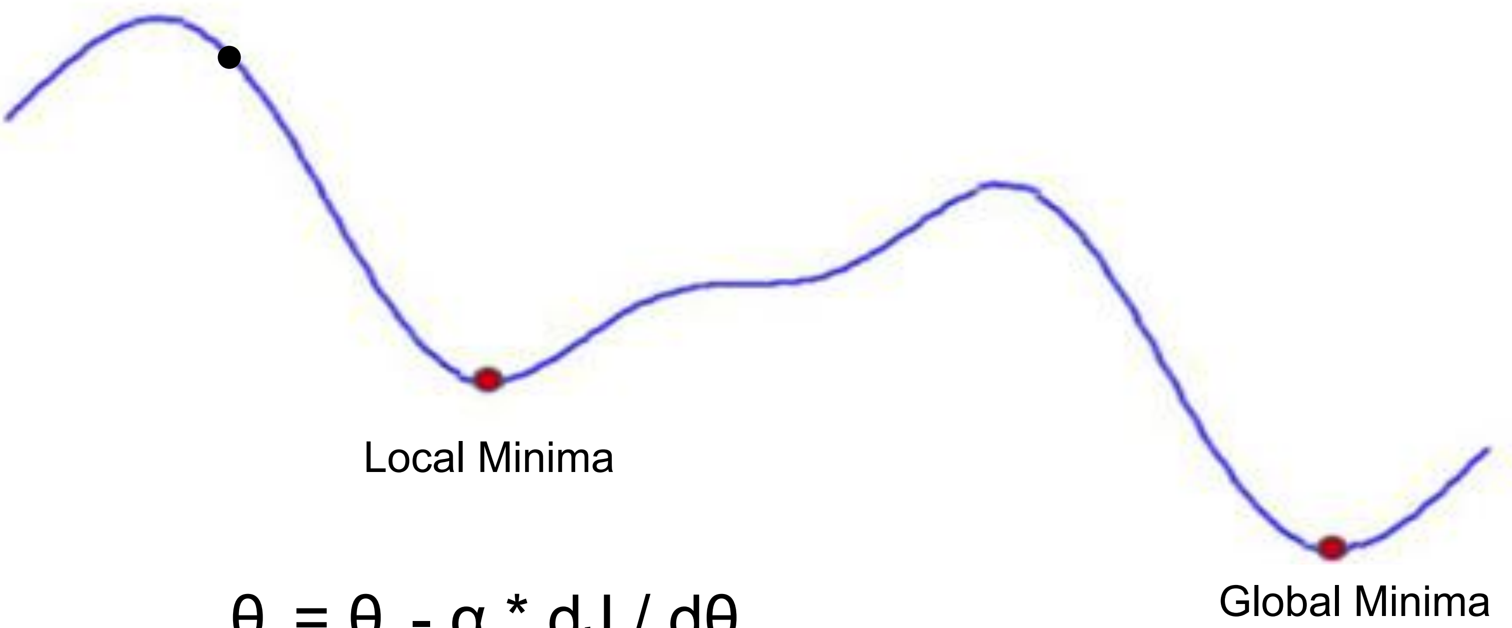
Problems with Gradient Descent

1. Getting stuck at local minima


Local Minima
Global Minima

Problems with Gradient Descent

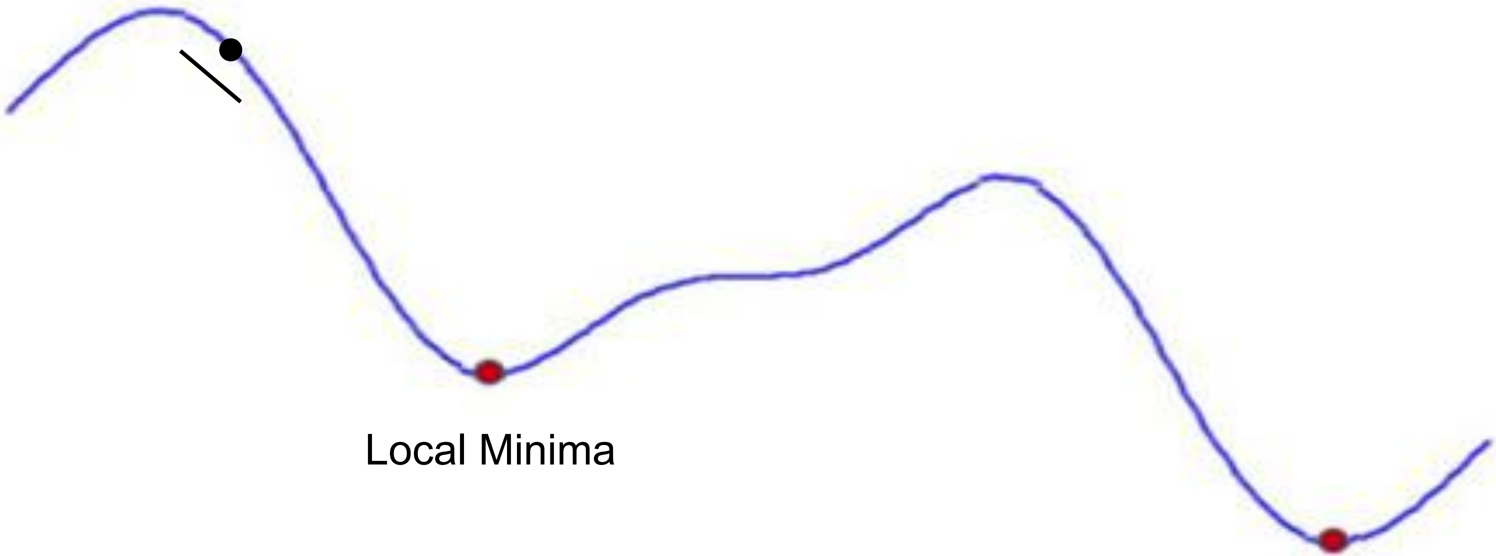
1. Getting stuck at local minima


Local Minima
Global Minima

$$\theta_i = \theta_i - \alpha * dJ / d\theta_i$$

Problems with Gradient Descent

1. Getting stuck at local minima



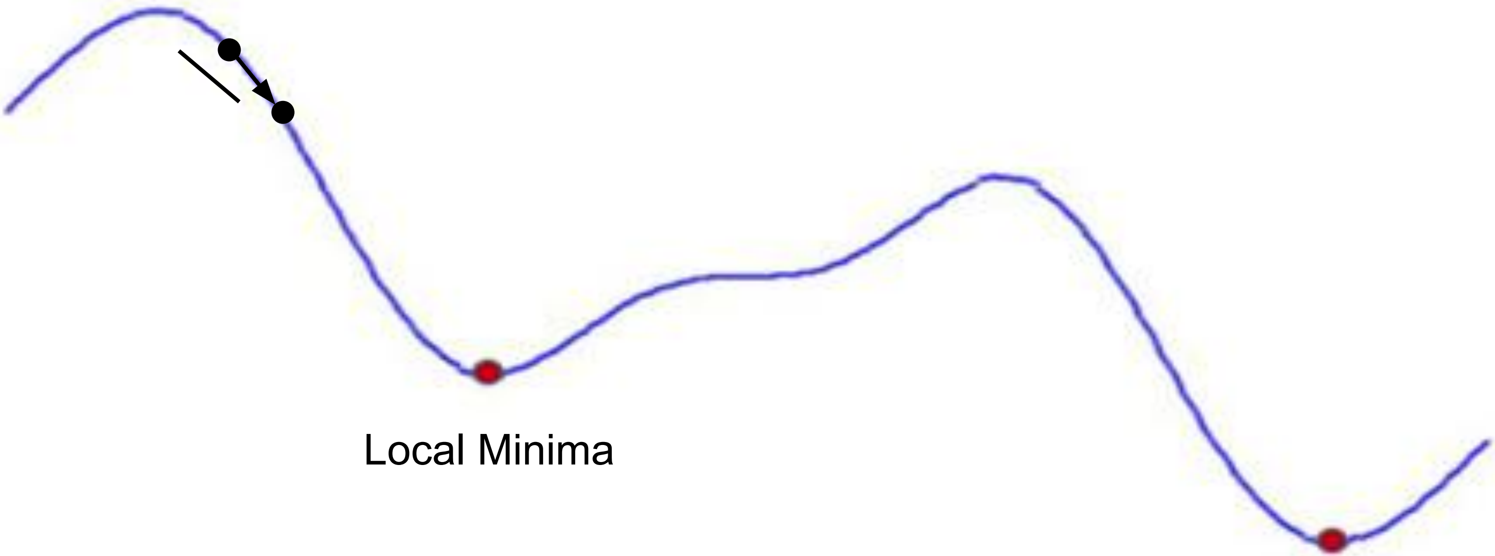
Local Minima

Global Minima

$$\theta_i = \theta_i - \alpha * dJ / d\theta_i$$

Problems with Gradient Descent

1. Getting stuck at local minima



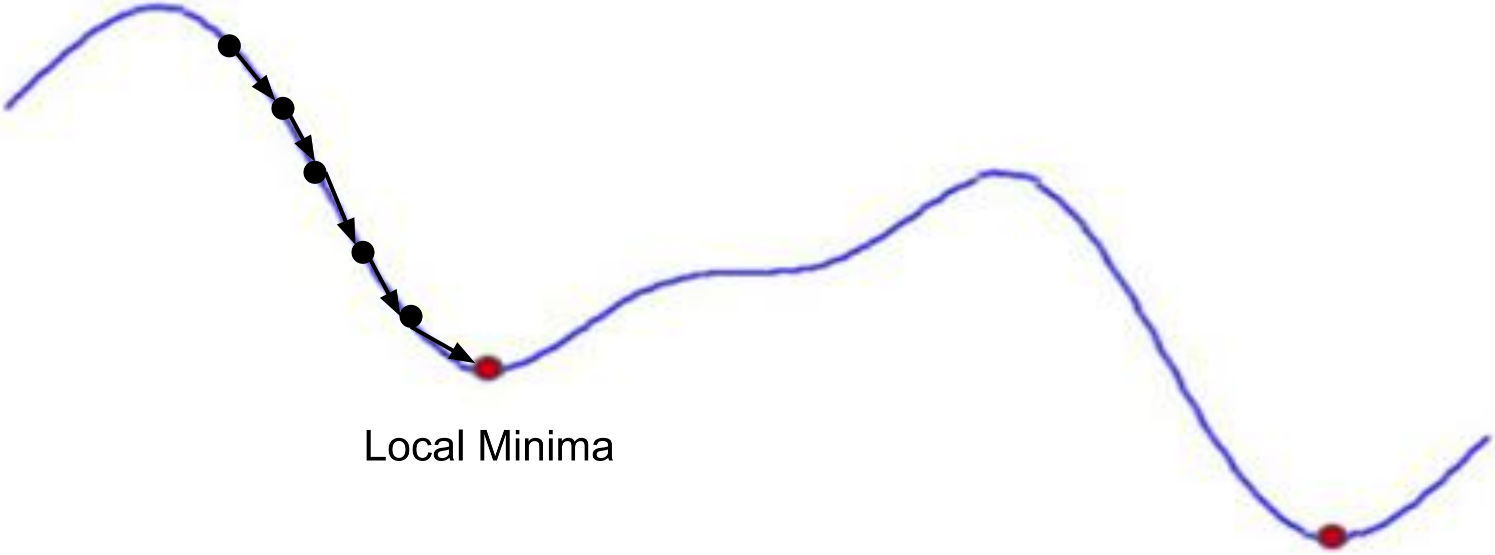
Local Minima

Global Minima

$$\theta_i = \theta_i - \alpha * dJ / d\theta_i$$

Problems with Gradient Descent

1. Getting stuck at local minima

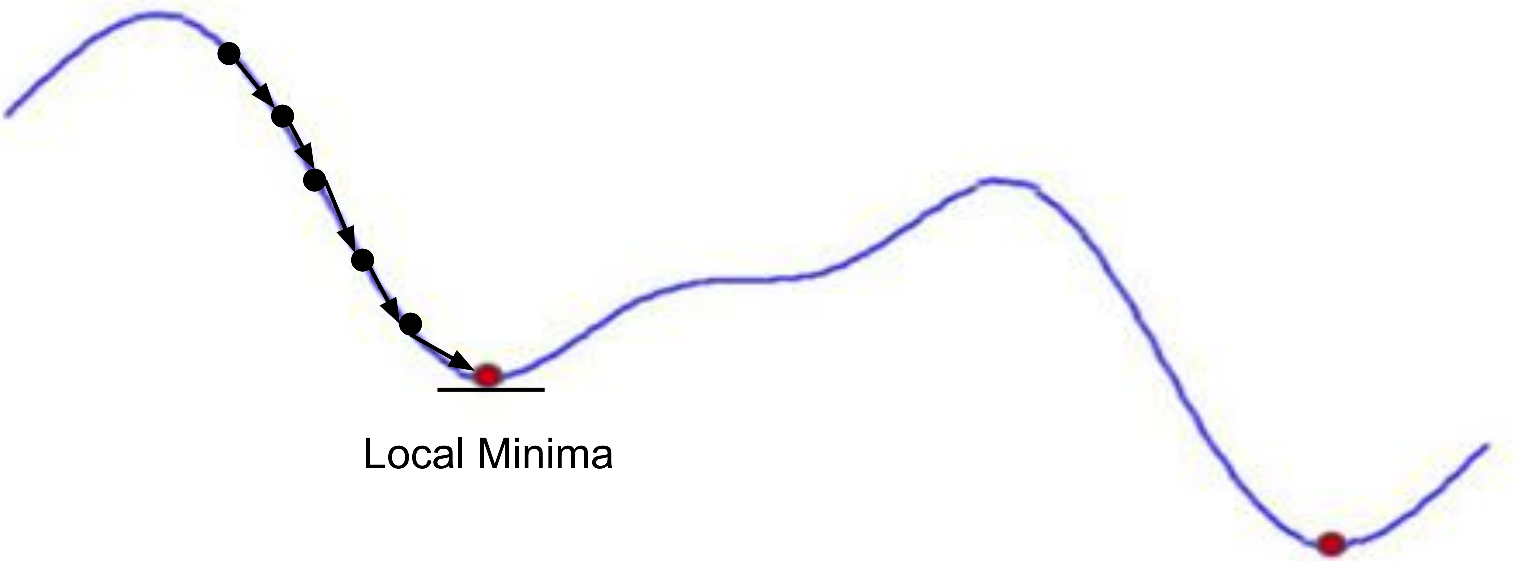


Local Minima
Global Minima

$$\theta_i = \theta_i - \alpha * dJ / d\theta_i$$

Problems with Gradient Descent

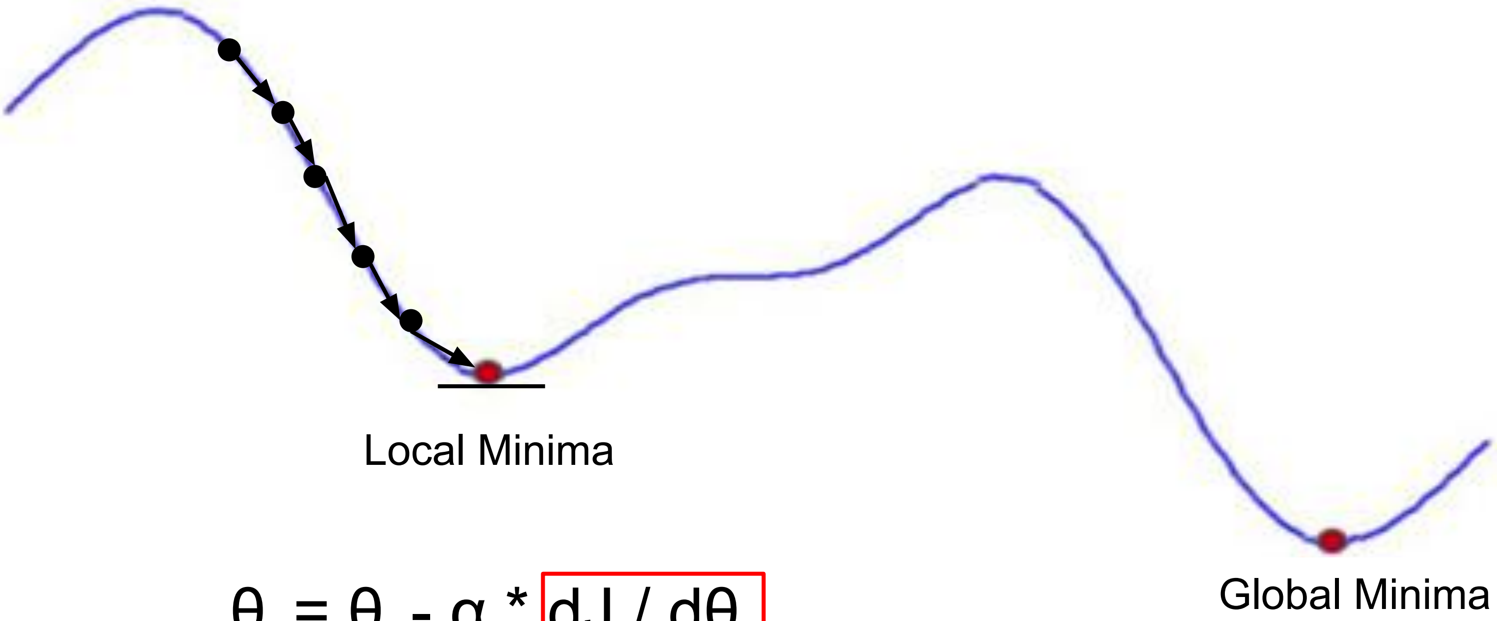
1. Getting stuck at local minima


Local Minima
Global Minima

$$\theta_i = \theta_i - \alpha * dJ / d\theta_i$$

Problems with Gradient Descent

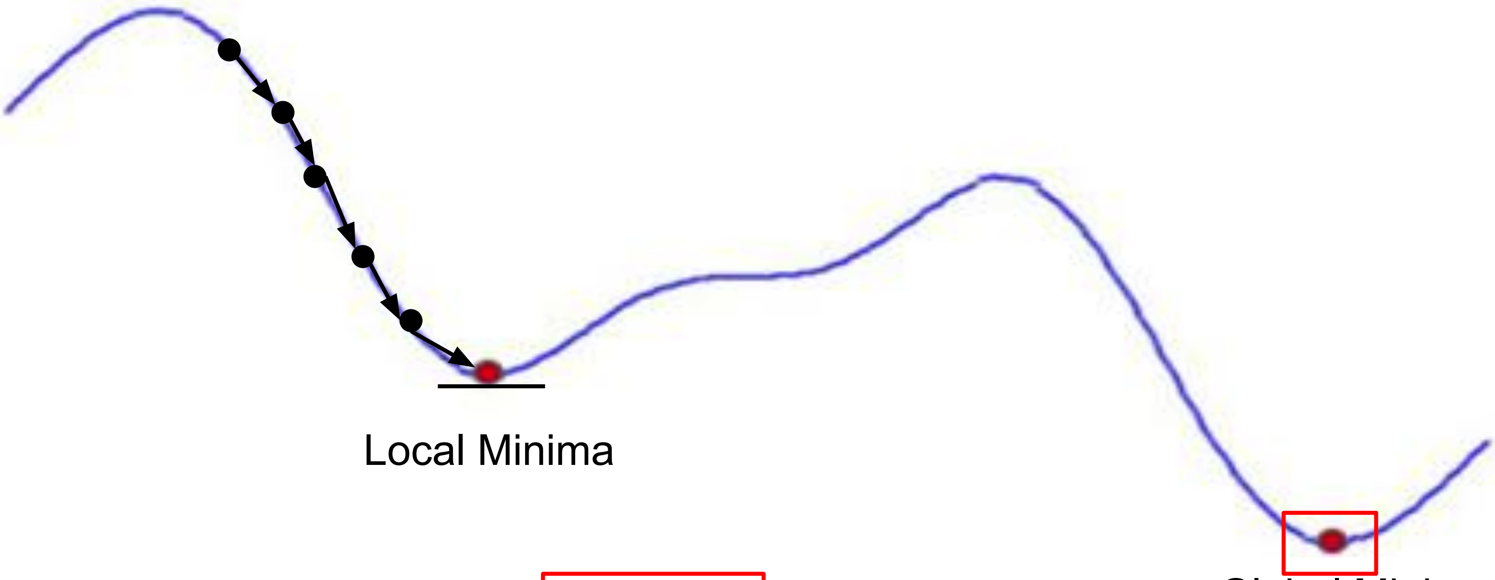
1. Getting stuck at local minima


Local Minima
Global Minima

$$\theta_i = \theta_i - \alpha * \boxed{dJ / d\theta_i}$$

Problems with Gradient Descent

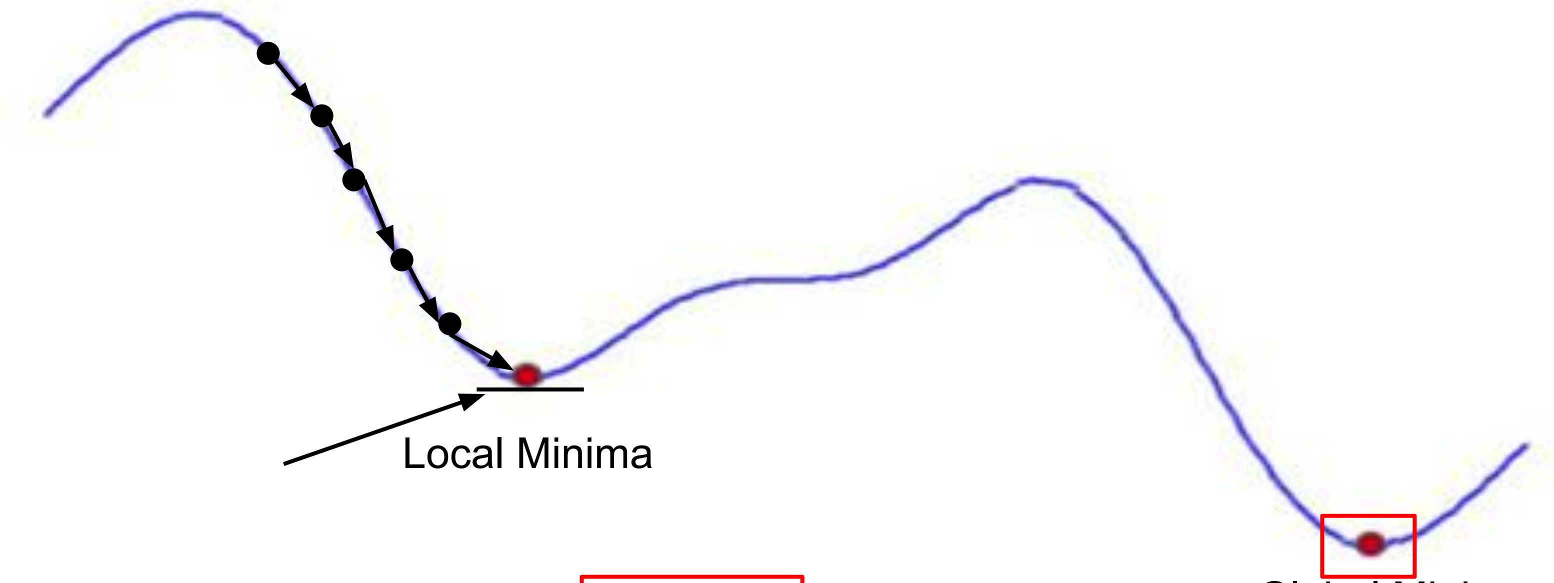
1. Getting stuck at local minima


Local Minima
Global Minima

$$\theta_i = \theta_i - \alpha * \boxed{dJ / d\theta_i}$$

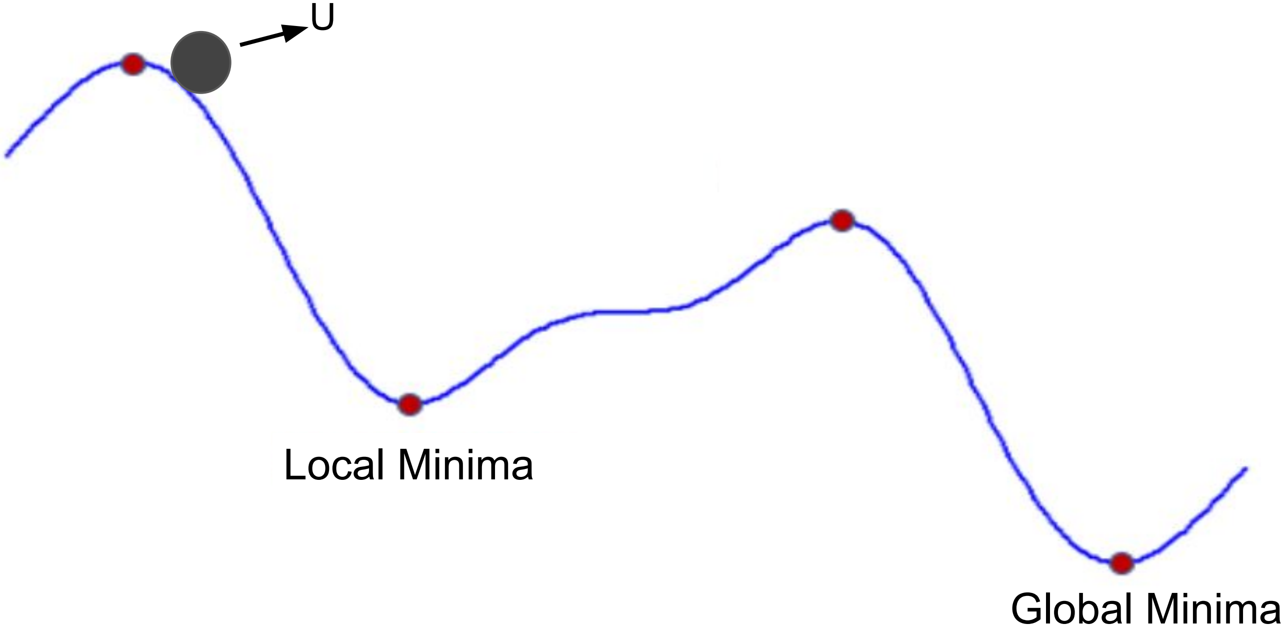
Problems with Gradient Descent

1. Getting stuck at local minima

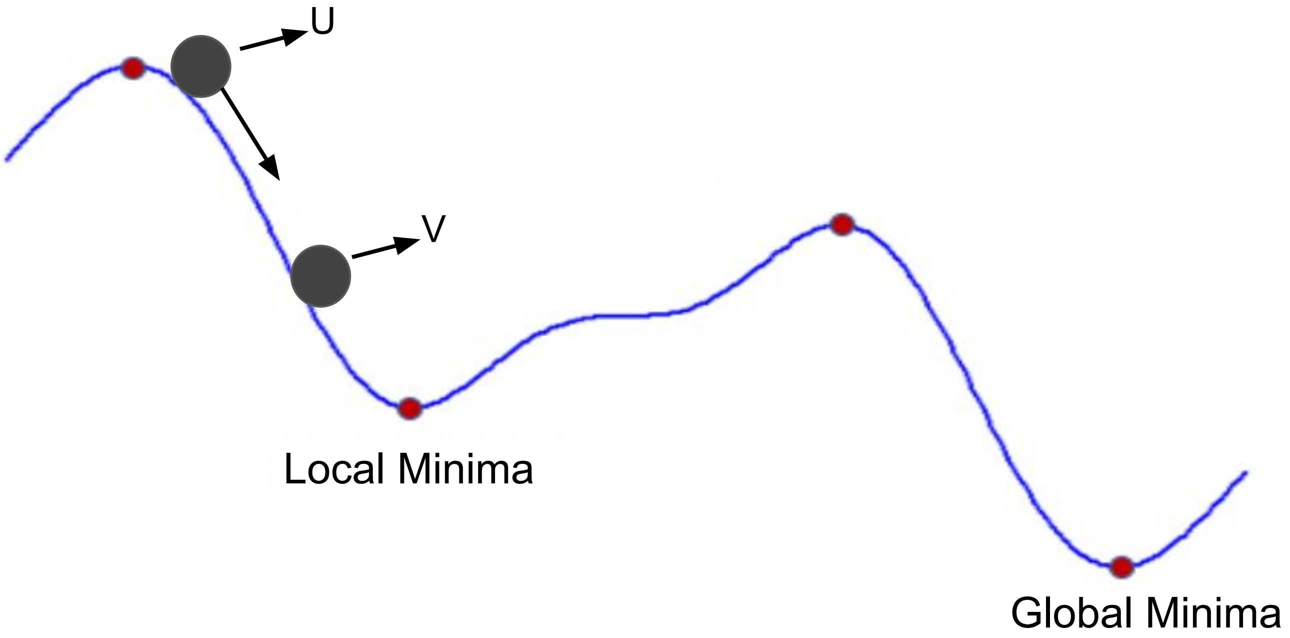

Local Minima
Global Minima

$$\theta_i = \theta_i - \alpha * \boxed{dJ / d\theta_i}$$

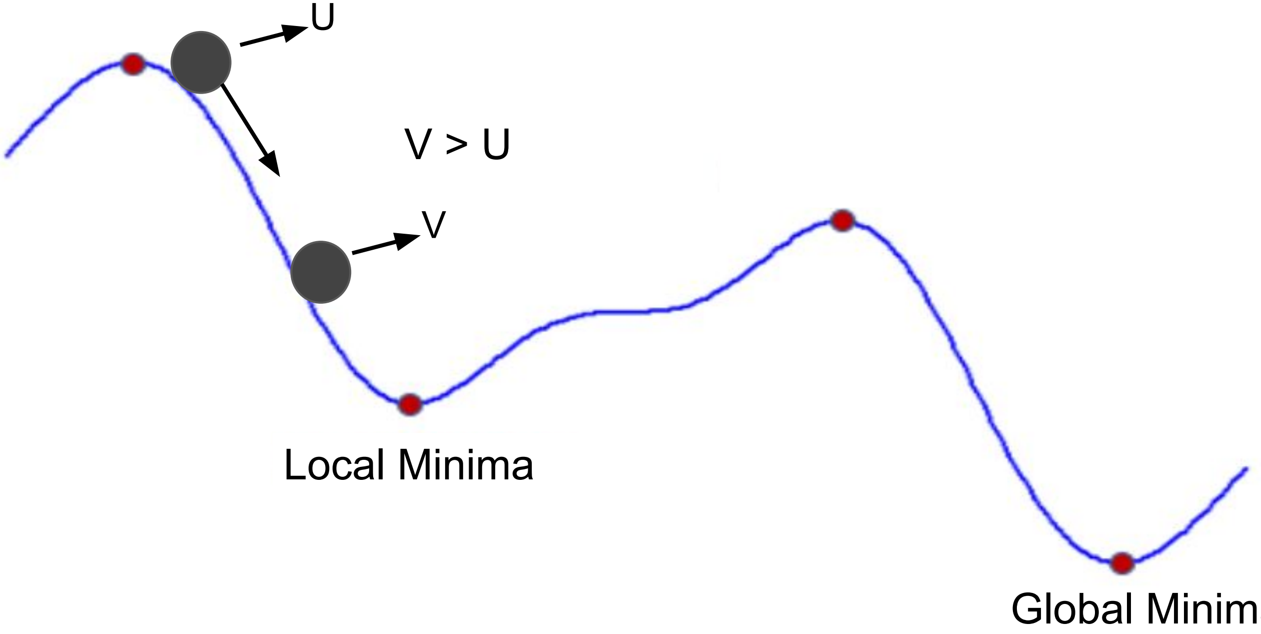
Solution: Getting stuck at local minima


U
Local Minima
Global Minima

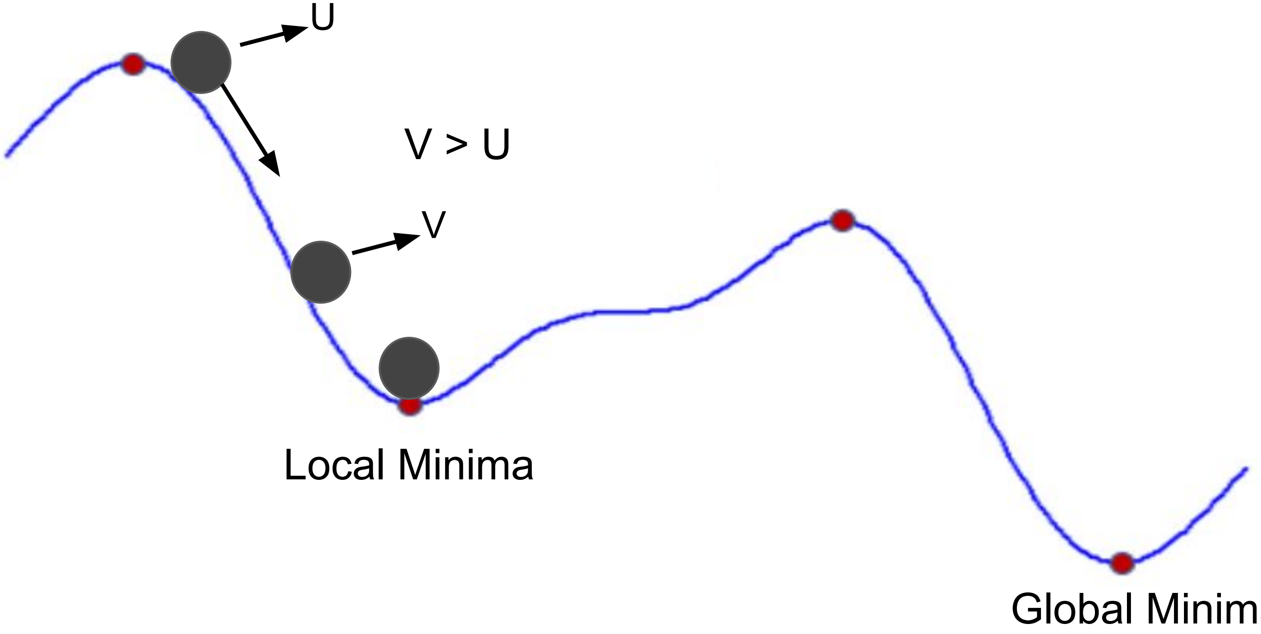
Solution: Getting stuck at local minima


U
V
Local Minima
Global Minima

Solution: Getting stuck at local minima

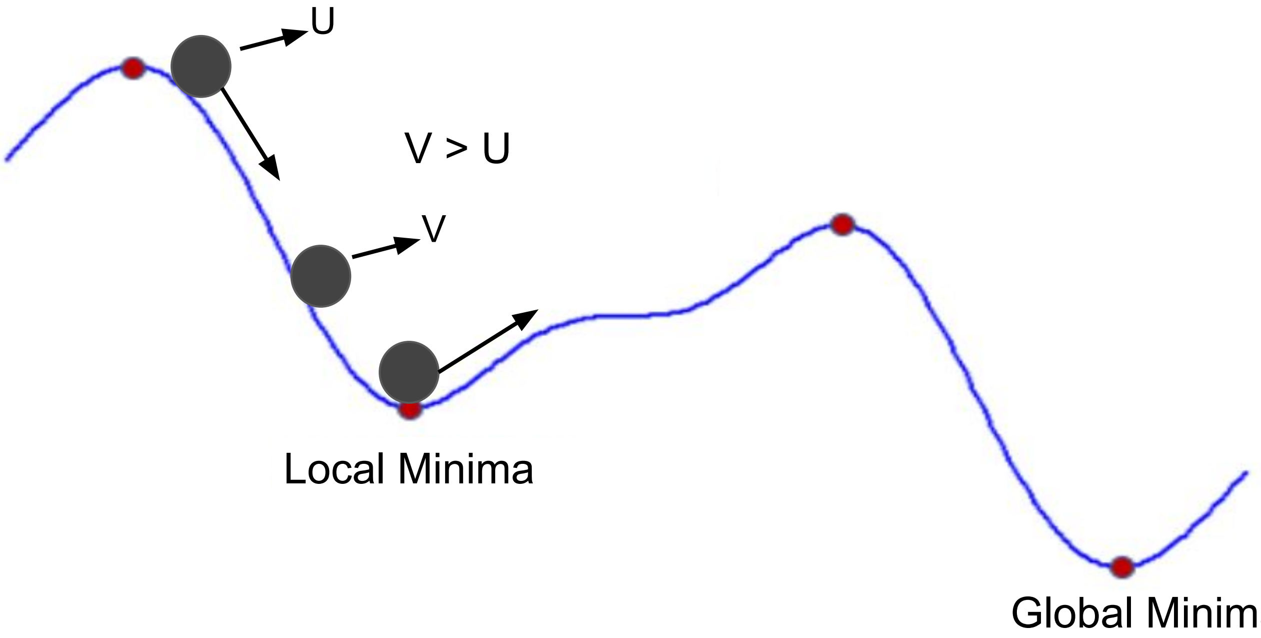

U
V > U
V
Local Minima
Global Minima

Solution: Getting stuck at local minima

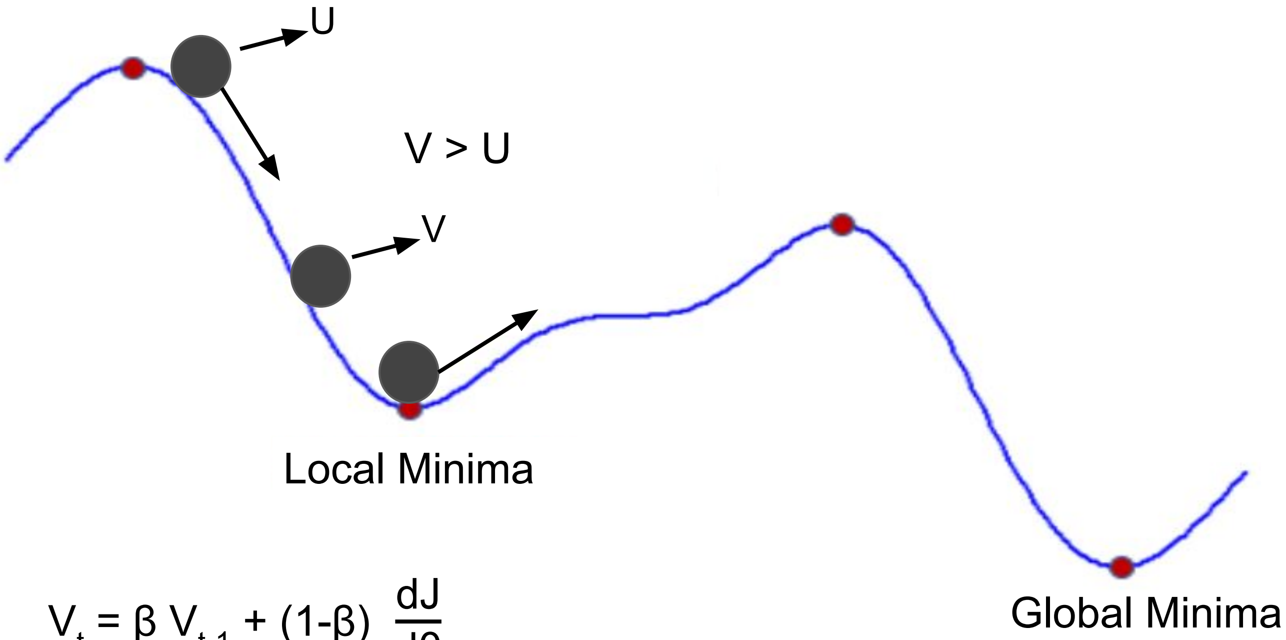

U
V > U
V
Local Minima
Global Minima

Analytics Vidhya
Learn everything about analytics

Solution: Getting stuck at local minima

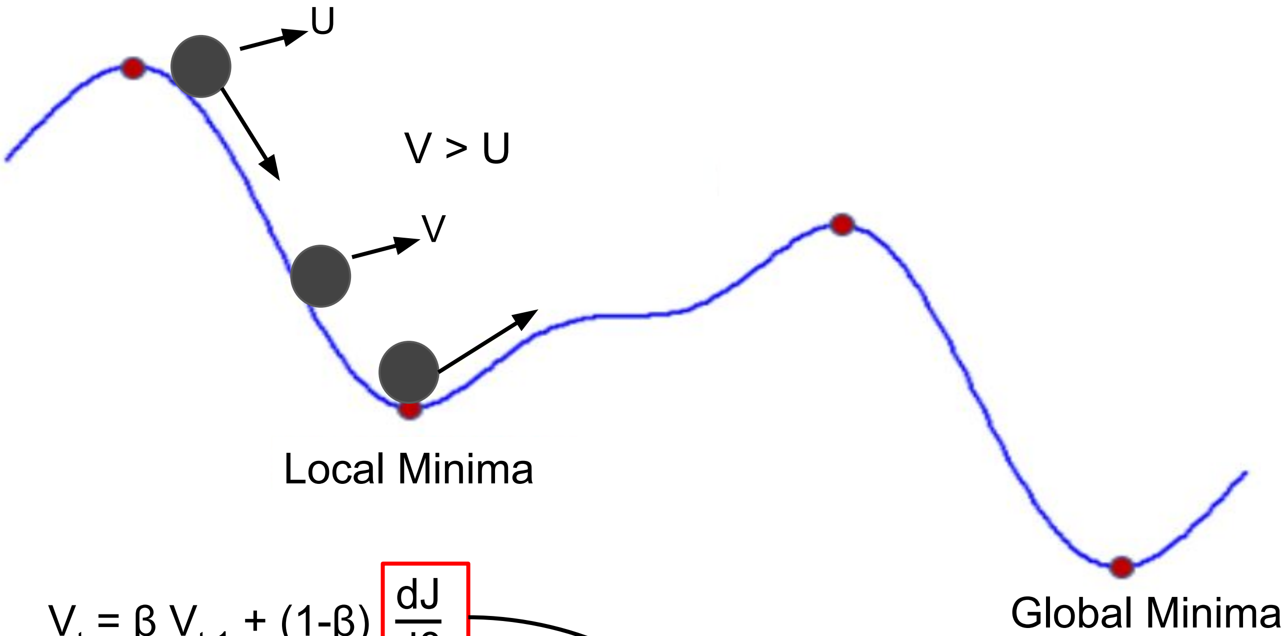

U
V > U
V
Local Minima
Global Minima

Weighted Gradient


U
V > U
V
Local Minima
Global Minima

$$V_t = \beta\, V_{t-1} + (1-\beta)\, \frac{dJ}{d\theta}$$

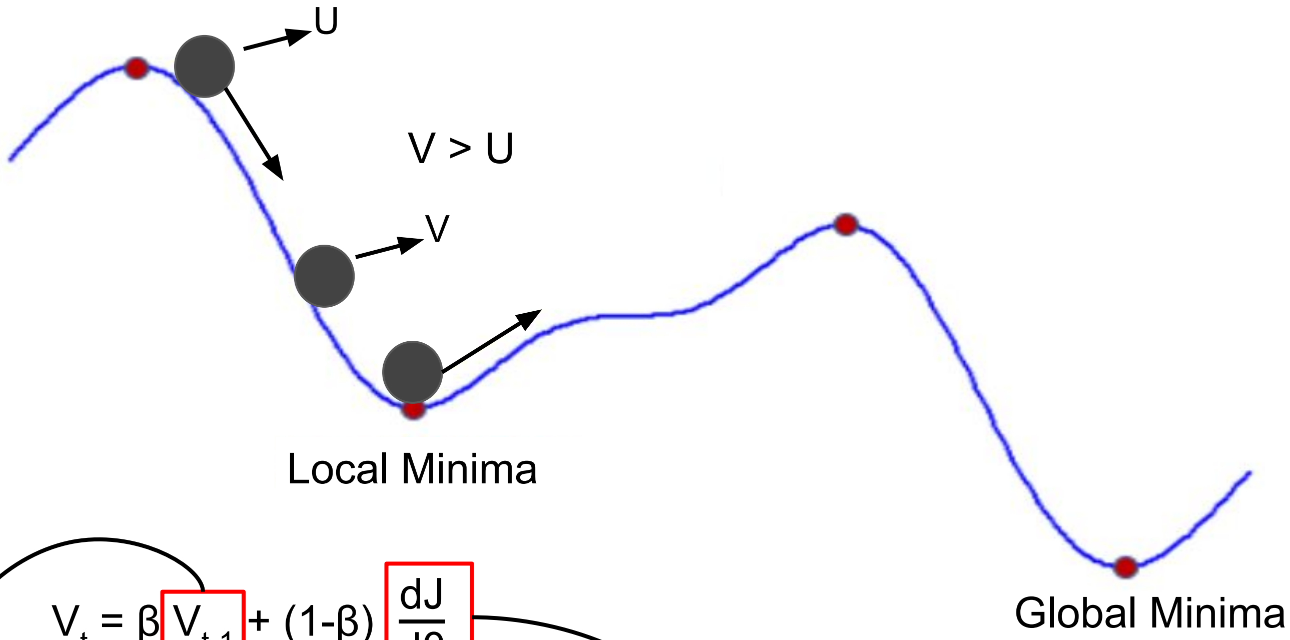
Weighted Gradient


U
V > U
V
Local Minima
Global Minima

$$V_t = \beta V_{t-1} + (1-\beta) \boxed{\frac{dJ}{d\theta}}$$

Current Gradient

Weighted Gradient

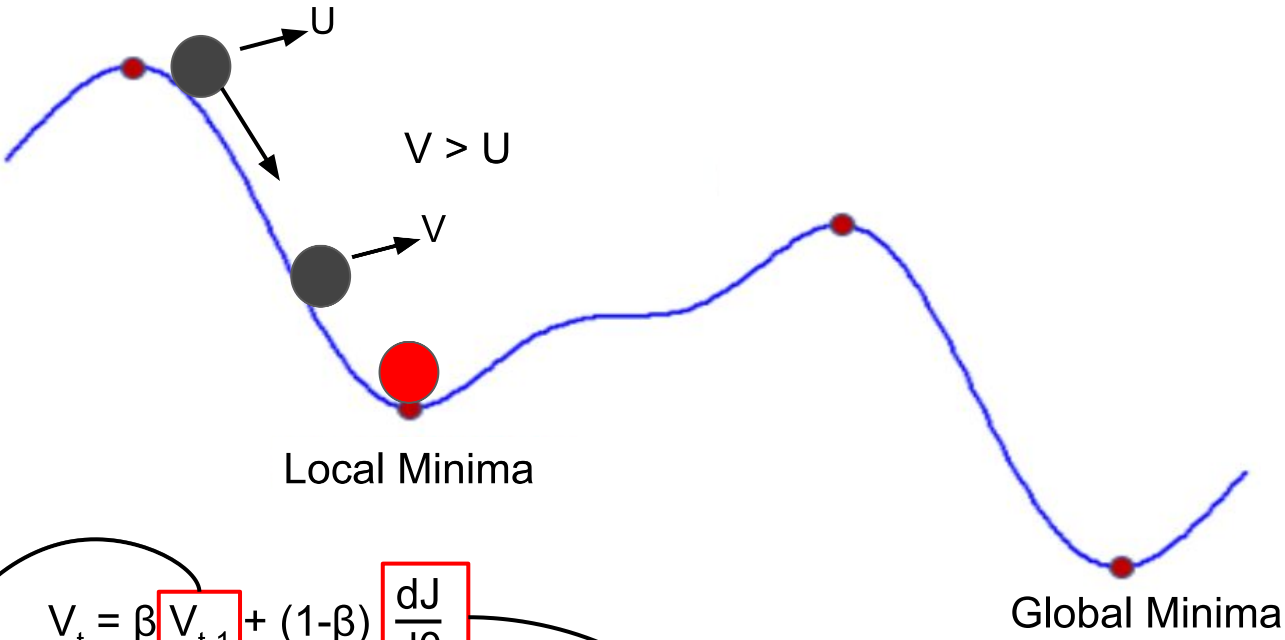

U
V > U
V
Local Minima
Global Minima

$$V_t = \beta \boxed{V_{t-1}} + (1-\beta) \boxed{\frac{dJ}{d\theta}}$$

Previously Accumulated Gradient

Current Gradient

Weighted Gradient

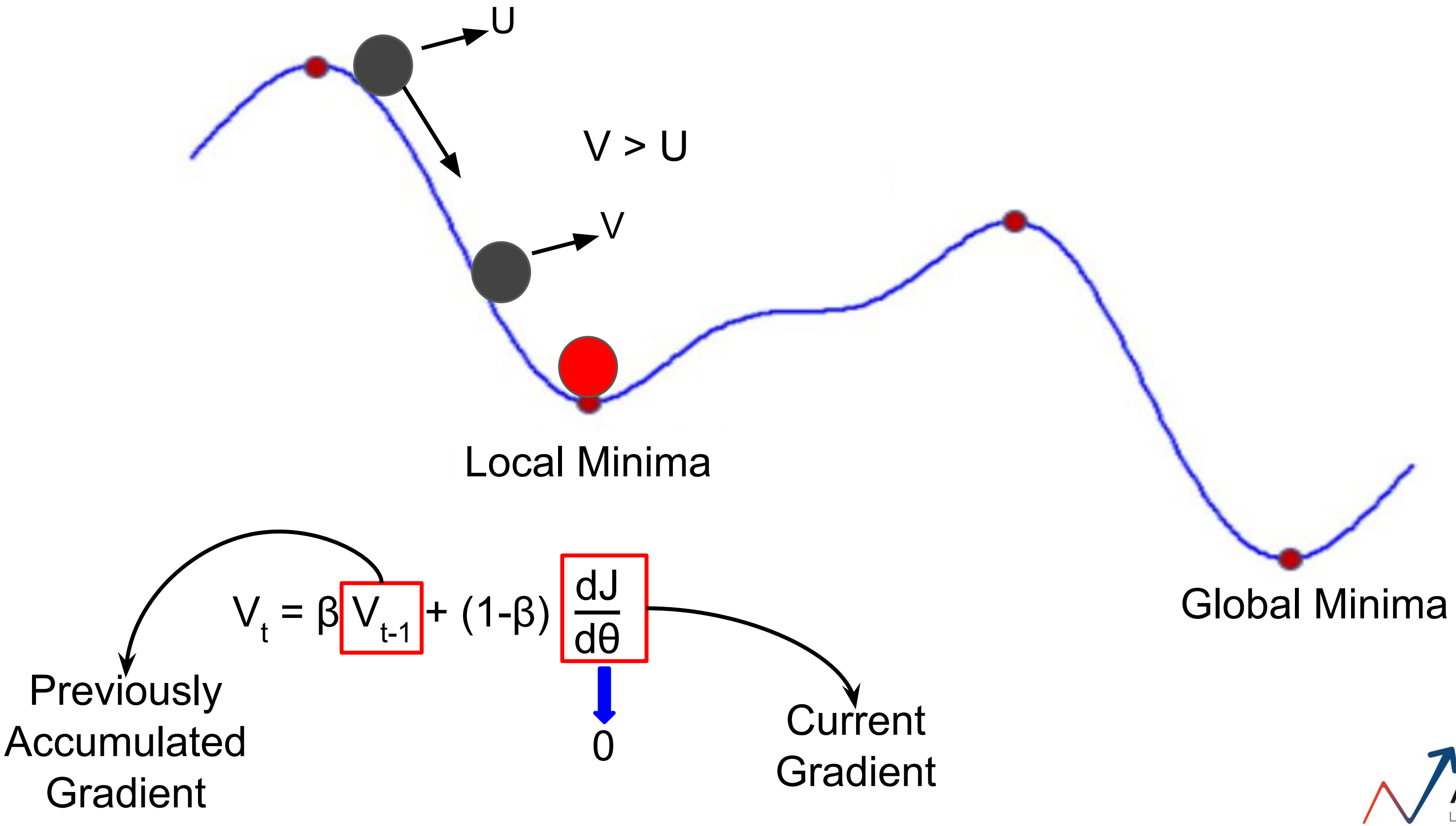

U
V > U
V
Local Minima
Global Minima

$$V_t = \beta \boxed{V_{t-1}} + (1-\beta) \boxed{\frac{dJ}{d\theta}}$$

Previously
Accumulated
Gradient

Current
Gradient

Weighted Gradient



U

V > U

V

Local Minima

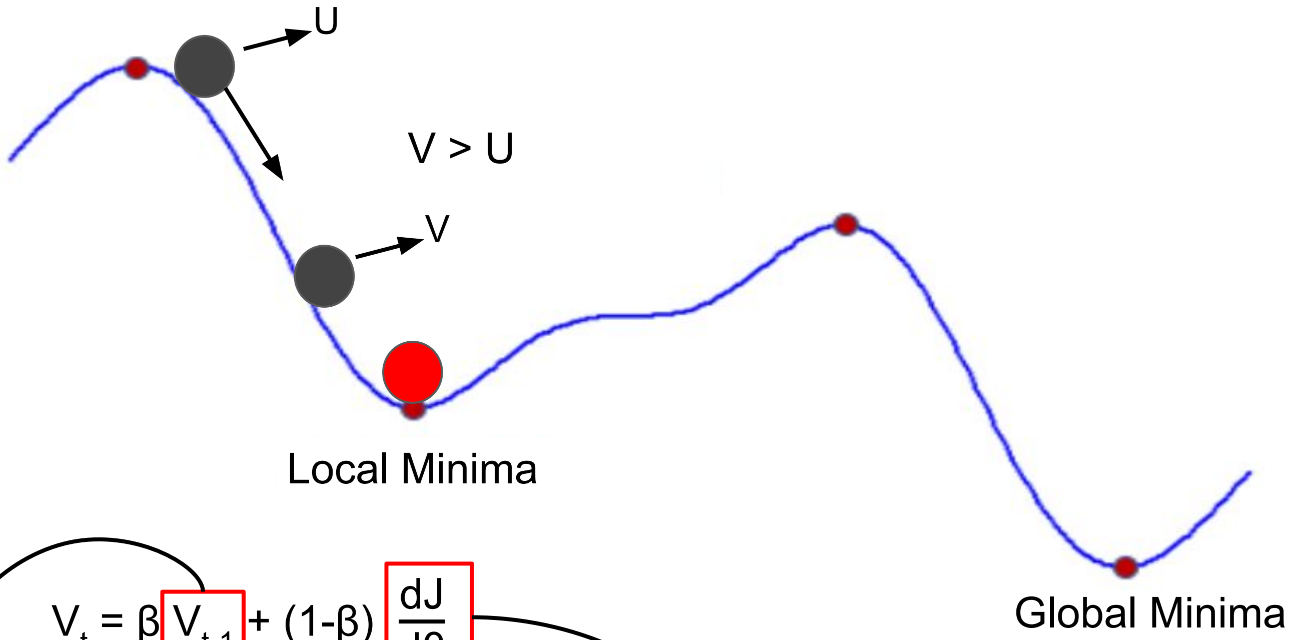
Global Minima

$$V_t = \beta \boxed{V_{t-1}} + (1-\beta) \underset{0}{\underset{\downarrow}{\boxed{\frac{dJ}{d\theta}}}}$$

Previously Accumulated Gradient

Current Gradient

Weighted Gradient


U
V > U
V
Local Minima
Global Minima

$$V_t = \beta \boxed{V_{t-1}} + (1-\beta) \boxed{\frac{dJ}{d\theta}}$$

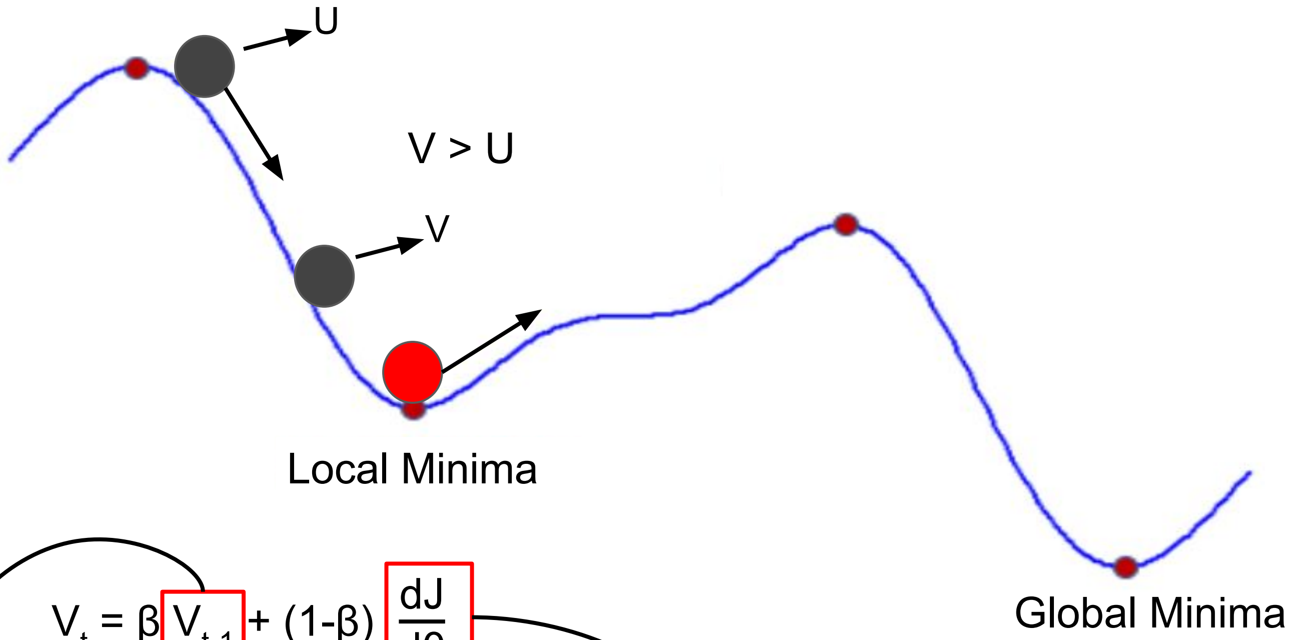
Previously Accumulated Gradient

Some value

0

Current Gradient

Weighted Gradient


U
V > U
V
Local Minima
Global Minima

$$V_t = \beta \boxed{V_{t-1}} + (1-\beta) \boxed{\frac{dJ}{d\theta}}$$

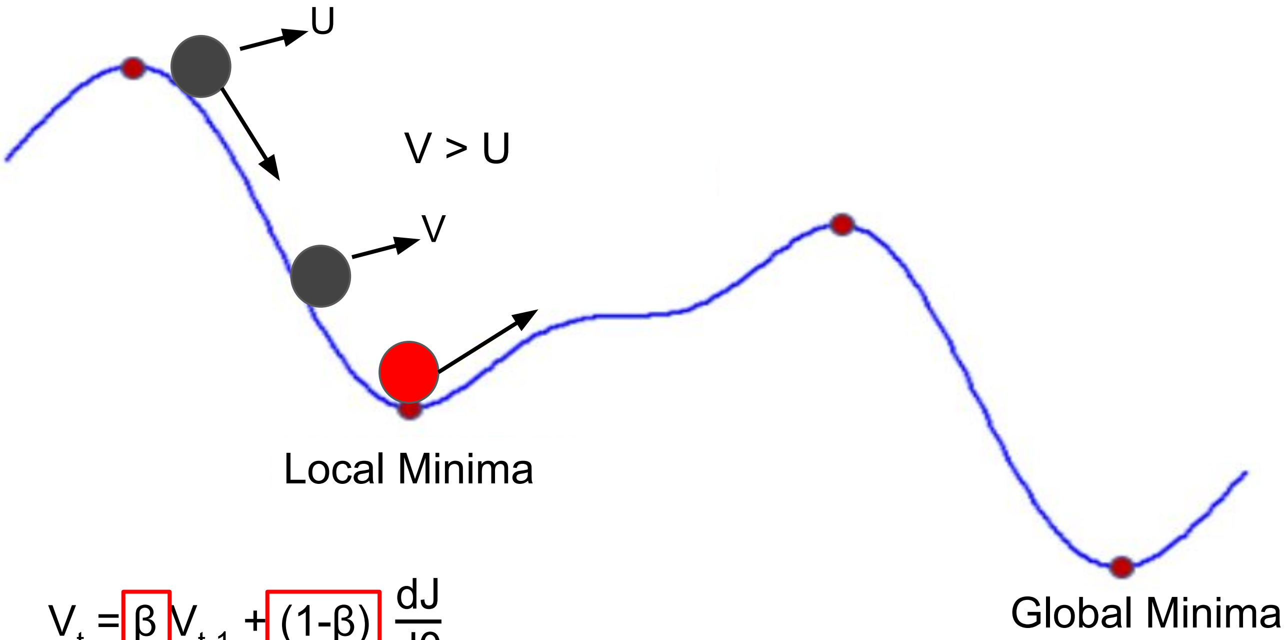
Previously Accumulated Gradient

Some value

0

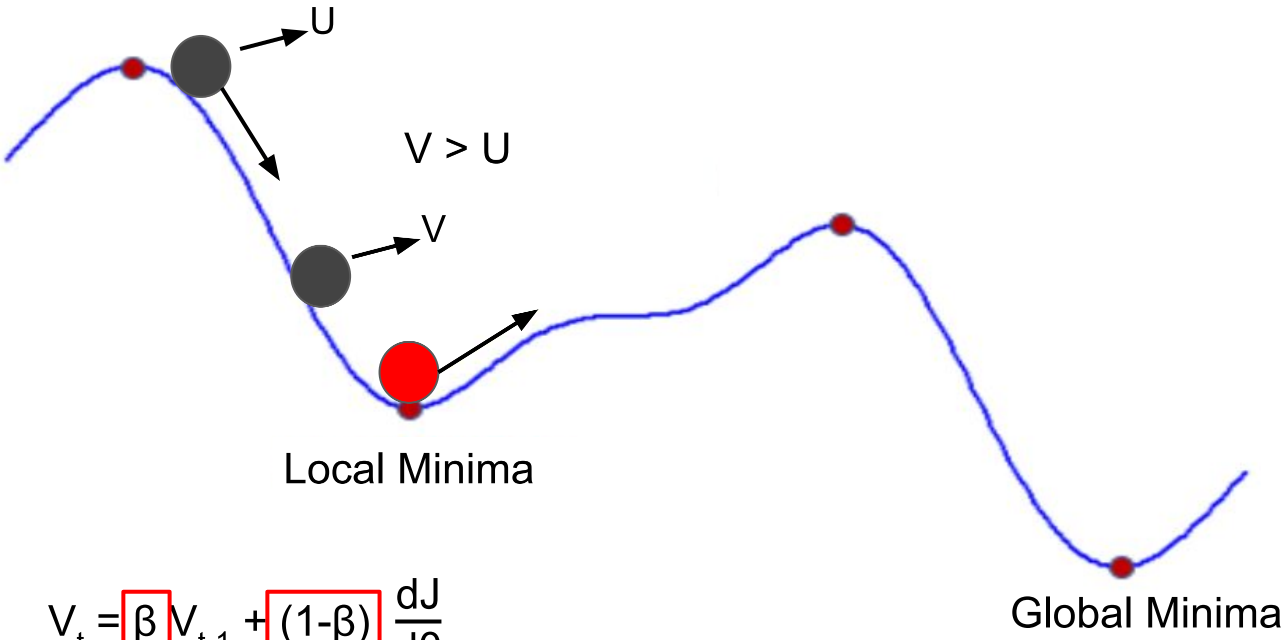
Current Gradient

Weighted Gradient


U
V > U
V
Local Minima
Global Minima

$$V_t = \boxed{\beta} V_{t-1} + \boxed{(1-\beta)} \frac{dJ}{d\theta}$$

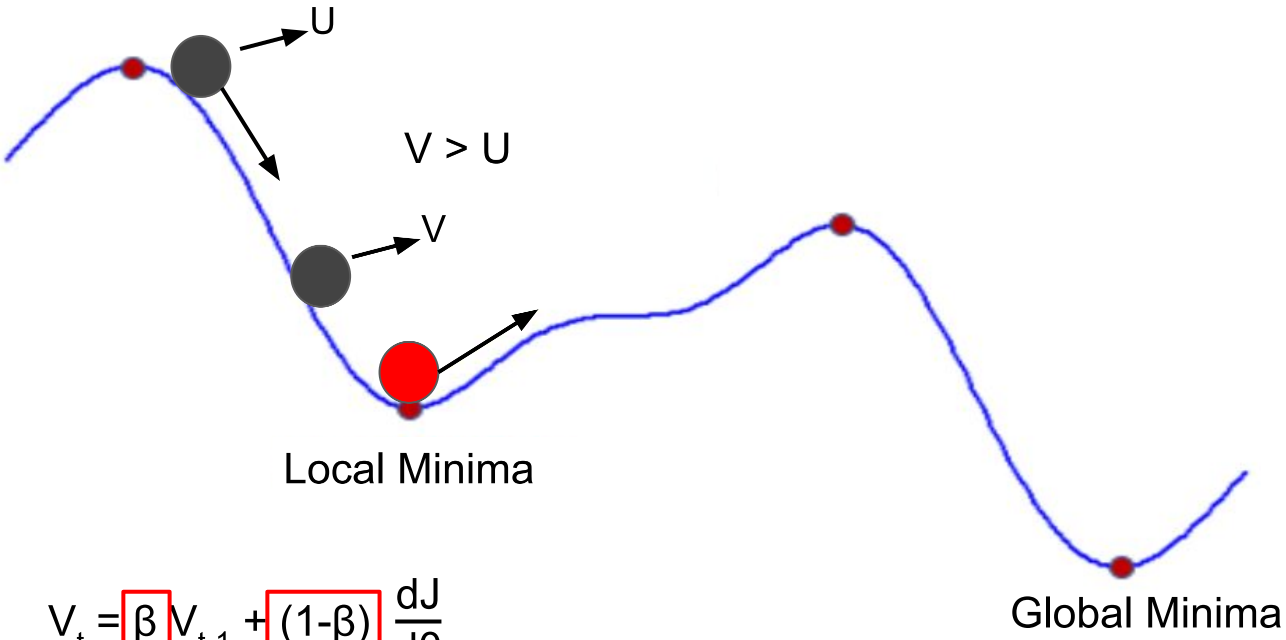
Weighted Gradient


U
V > U
V
Local Minima
Global Minima

$$V_t = \boxed{\beta} V_{t-1} + \boxed{(1-\beta)} \frac{dJ}{d\theta}$$

$\beta = 0.9$

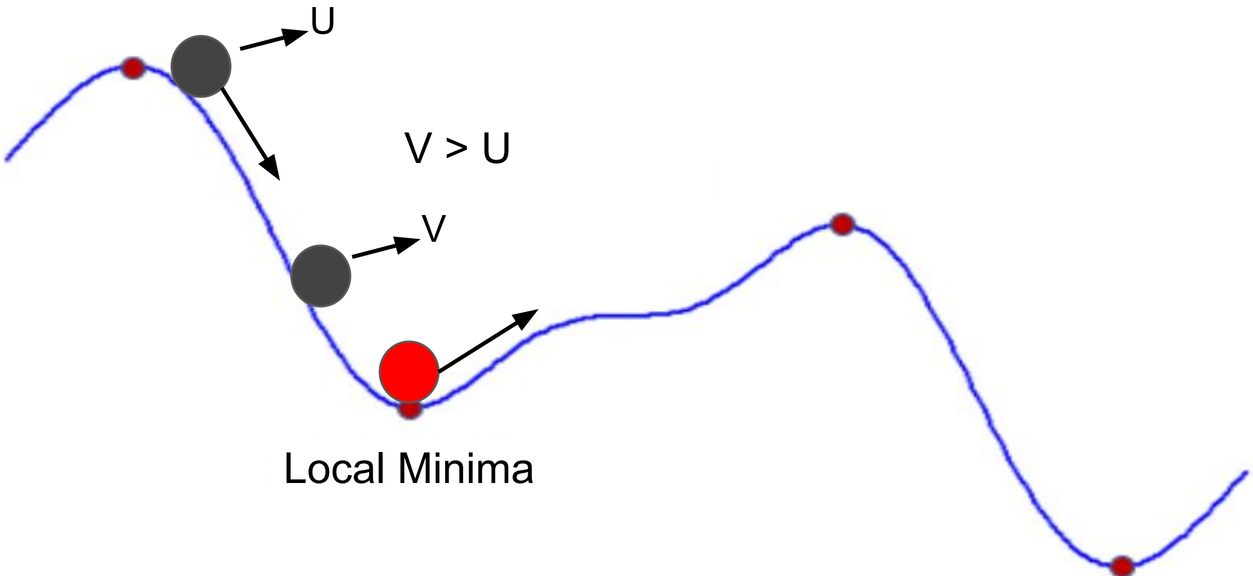
Weighted Gradient


U
V > U
V
Local Minima
Global Minima

$$V_t = \boxed{\beta} V_{t-1} + \boxed{(1-\beta)} \frac{dJ}{d\theta}$$

$\beta = 0.9$ → 10% weight to current gradient

Weighted Gradient


U
V > U
V
Local Minima
Global Minima

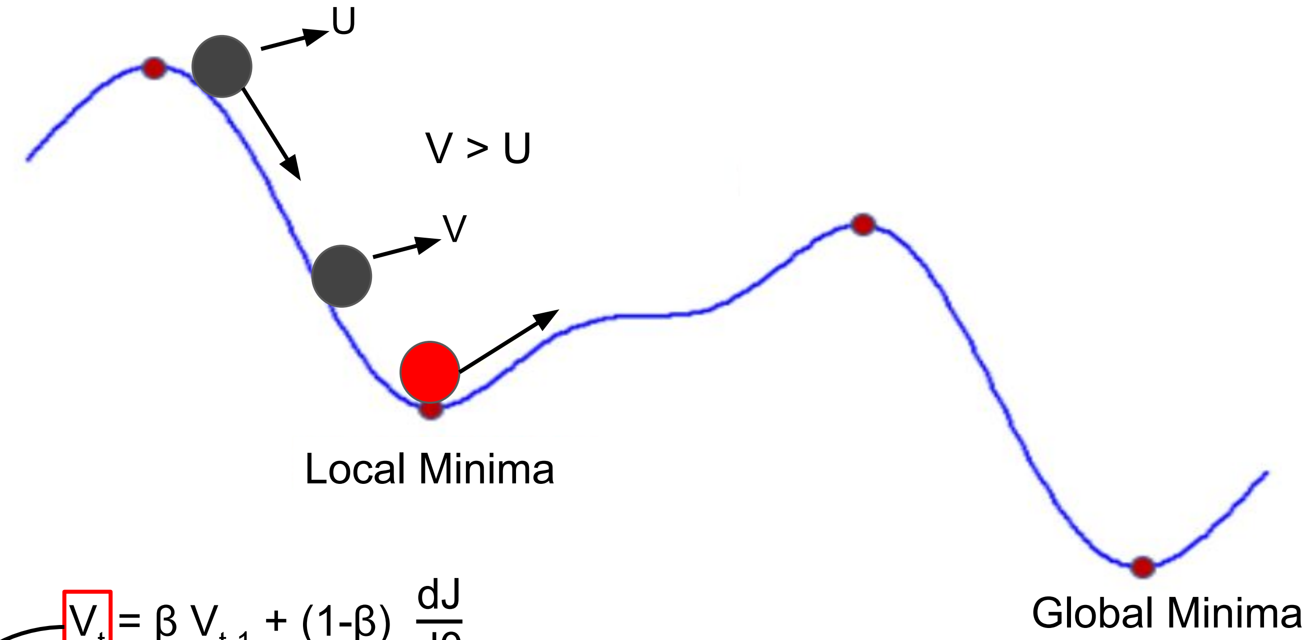
$$V_t = \boxed{\beta} V_{t-1} + \boxed{(1-\beta)} \frac{dJ}{d\theta}$$

$\beta = 0.9$

10% weight to current gradient

90% weight to previous gradient

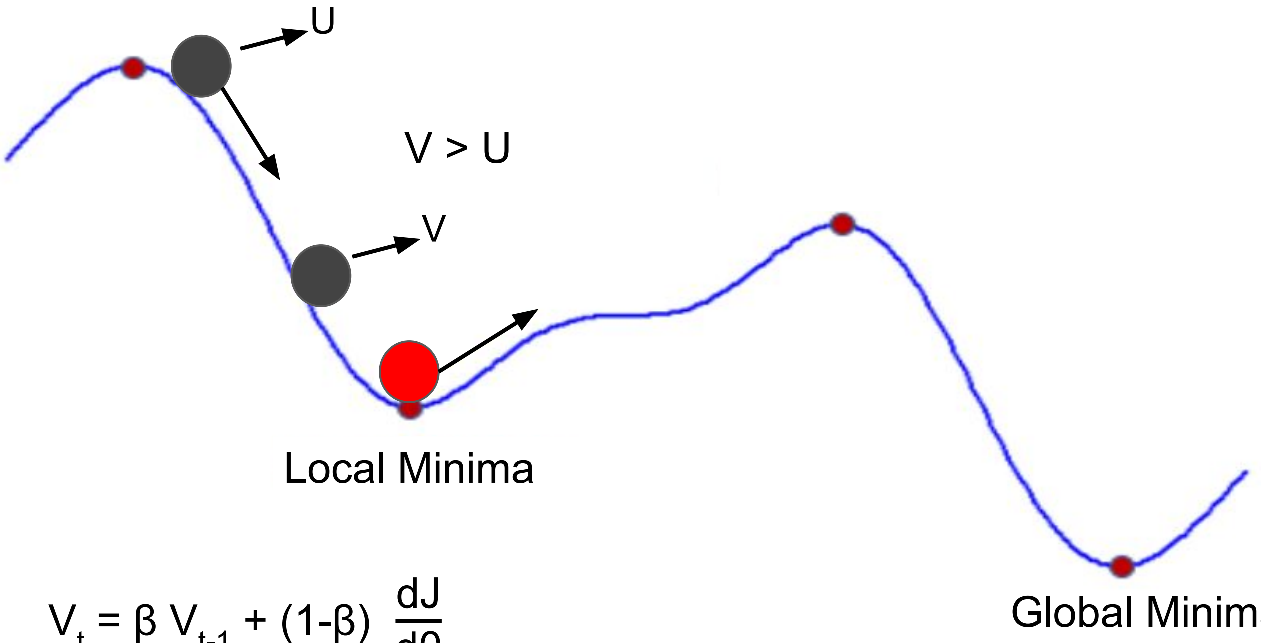
Weighted Gradient


U
V > U
V
Local Minima
Global Minima

$$\boxed{V_t} = \beta V_{t-1} + (1-\beta)\frac{dJ}{d\theta}$$

New Gradient

Stochastic Gradient Descent with Momentum


U
V > U
V
Local Minima
Global Minima

$$V_t = \beta\, V_{t-1} + (1-\beta)\, \frac{dJ}{d\theta}$$

$$\theta_i = \theta_i - \alpha * V_t$$

Analytics Vidhya
Learn everything about analytics

Analytics
Vidhya

Thank You