

Understanding Gradients



$$\text{Loss} = Y_{\text{predicted}} - Y_{\text{actual}}$$



$$\text{Loss} = Y_{\text{predicted}} - Y_{\text{actual}}$$

Error in the prediction for a single row of data



$$\text{Loss} = 0$$



$$\text{Loss} = \Theta$$

For a single neuron,

$$\text{Gradient}(\Theta) = (\partial\Theta/\partial W_1, \partial\Theta/\partial b_1)$$



$$\text{Loss} = \Theta$$

For a network,

Gradient: Tensor of all partial derivatives



$$\text{Loss} = \Theta$$

Gradient: $(\partial\Theta/\partial W_1, \partial\Theta/\partial b_1, \partial\Theta/\partial W_2, \partial\Theta/\partial b_2, \dots, \partial\Theta/\partial W_M, \partial\Theta/\partial b_M)$



There are many ways of calculating these
gradients & PyTorch uses a way called
Automatic Differentiation



This is achieved using the library called
Autograd



Gradient(θ^t)



Gradient(θ^t)

Calculated at a specific instance t

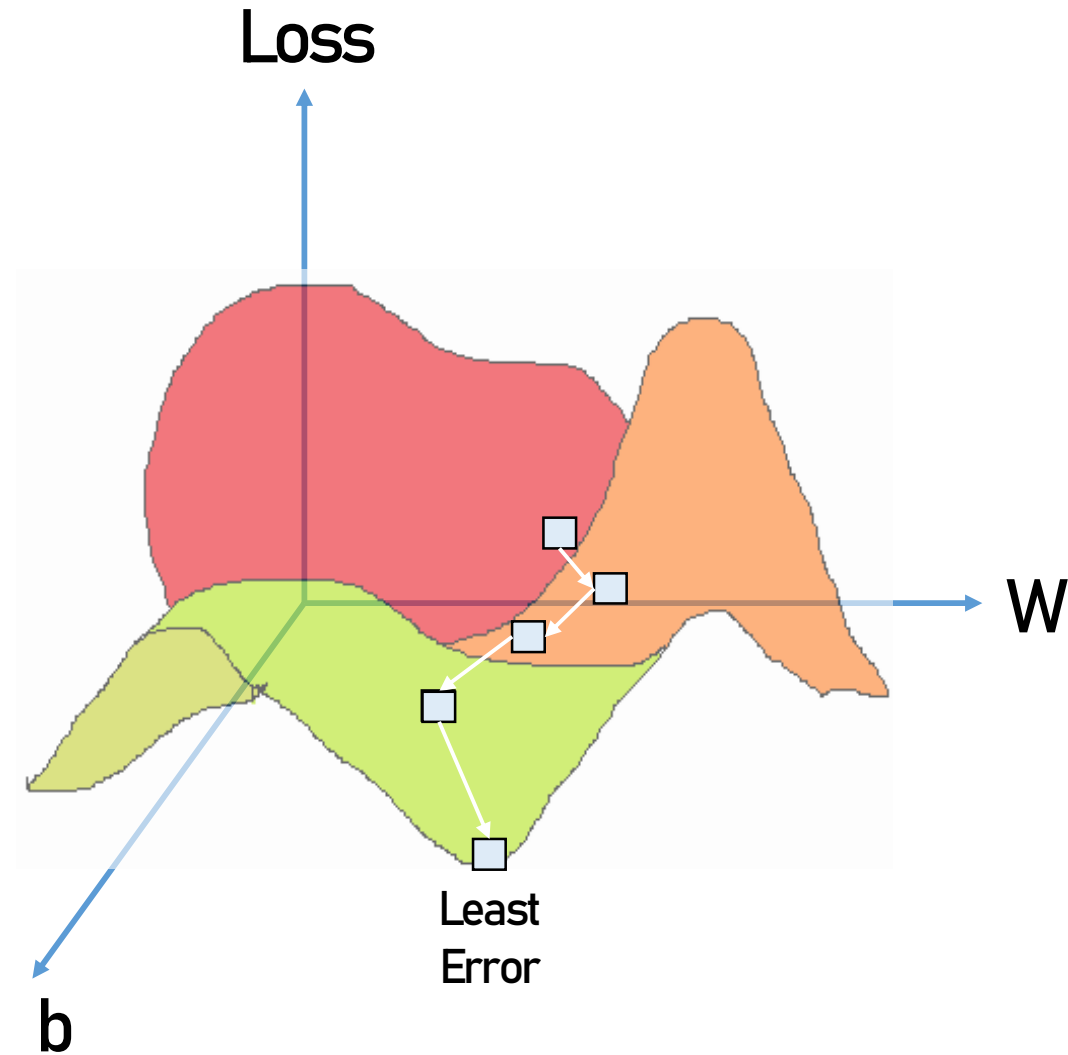


$$\text{Parameters}^{t+1} = \text{Parameters}^t - \text{learning_rate} * \text{Gradient}(\theta^t)$$

Change each parameter value by deducting the respective gradient multiplied by the learning rate



Visualizing Gradient Descent

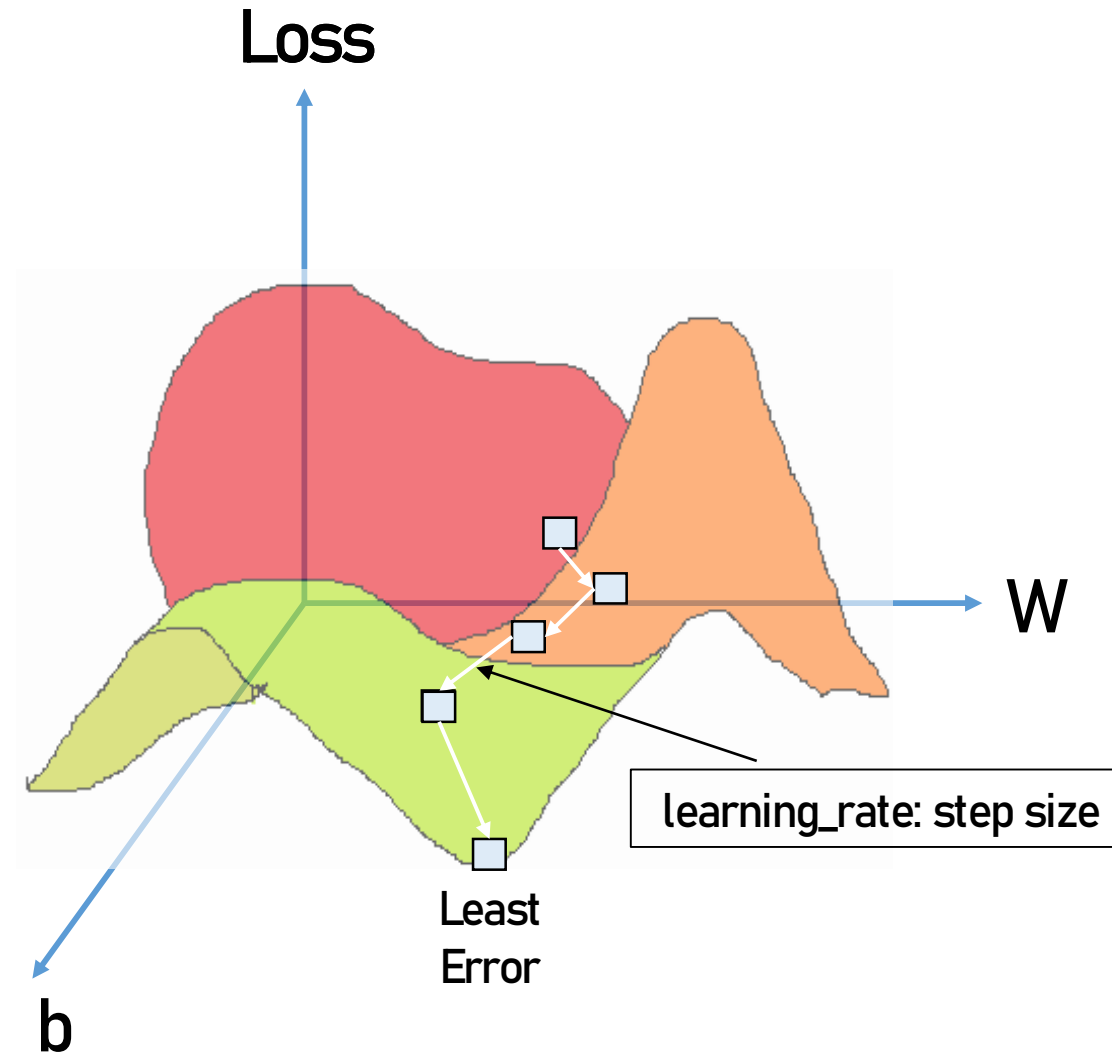


$$\text{Parameters}^{t+1} = \text{Parameters}^t - \text{learning_rate} * \text{Gradient}(\theta^t)$$

Change each parameter value by deducting the respective gradient multiplied by the learning rate



Visualizing Gradient Descent



Choosing the value of learning rate

If learning rate is too small then parameters take a very long time to converge



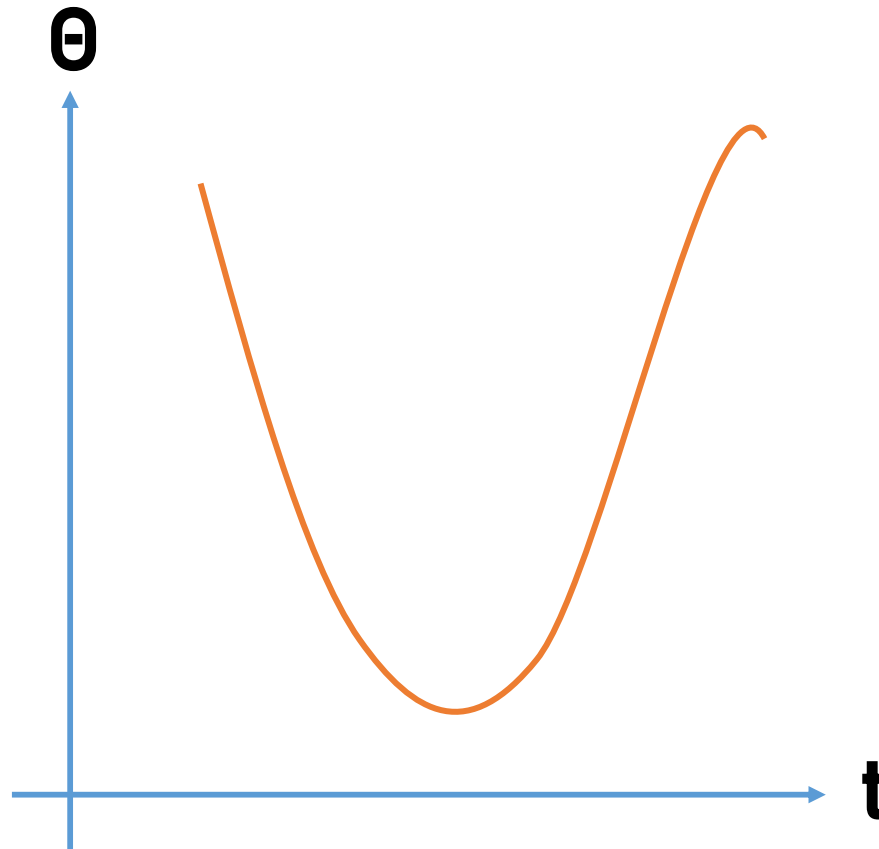
Choosing the value of learning rate

If learning rate is too small then parameters take a very long time to converge

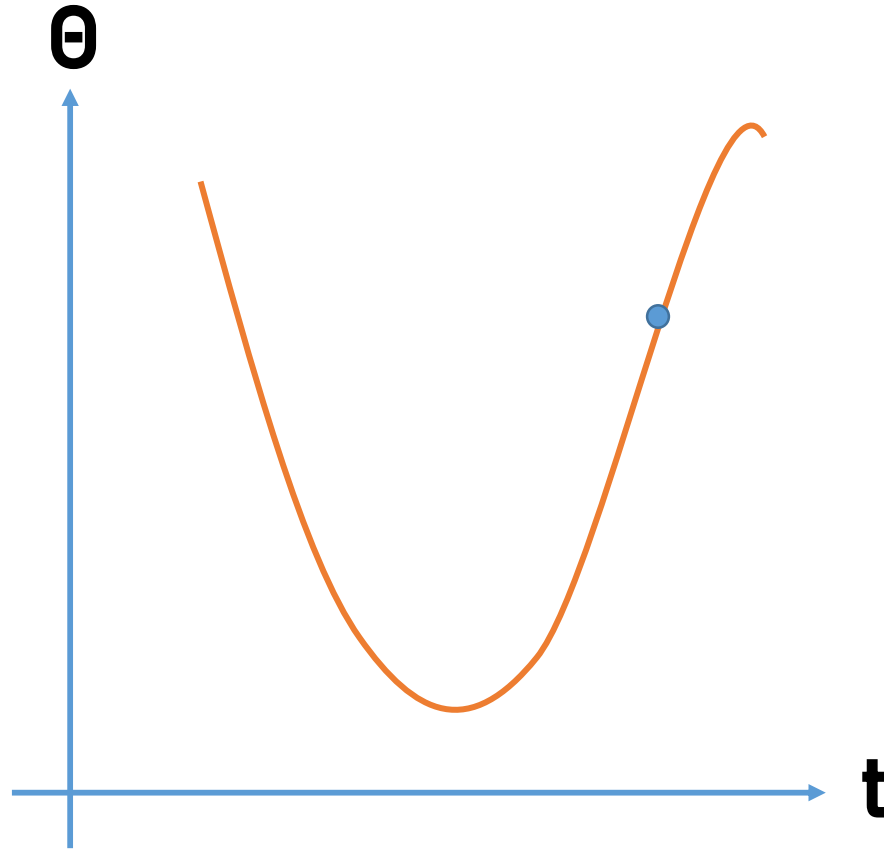
If learning rate is too big then there's a chance that parameter values will explode and never converge



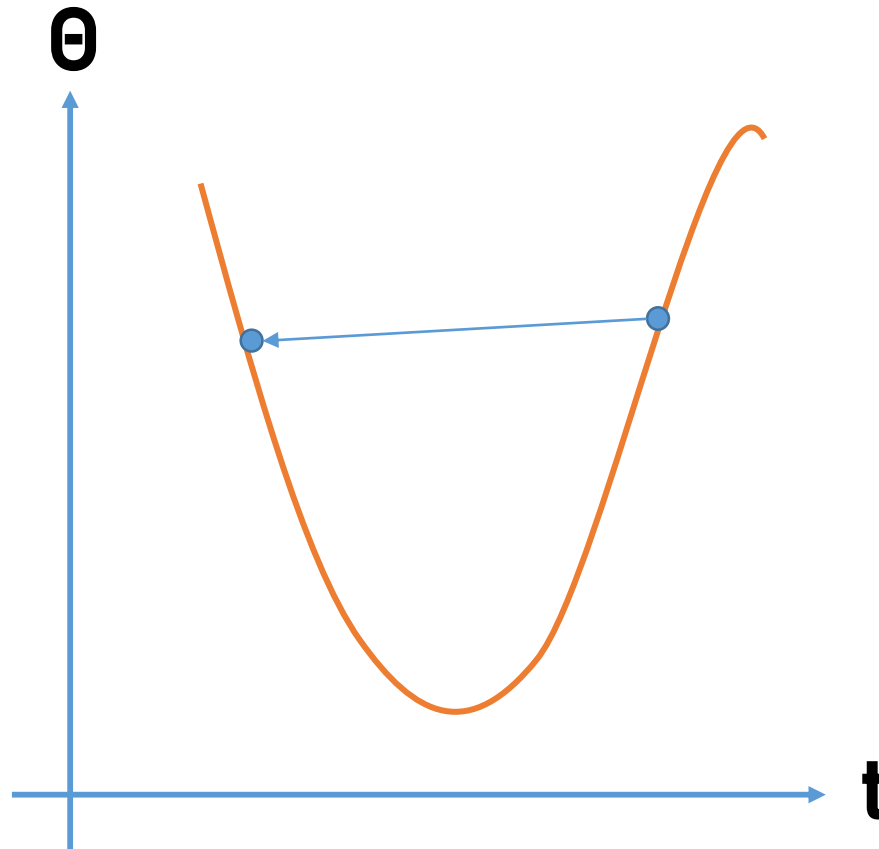
Choosing the value of learning rate



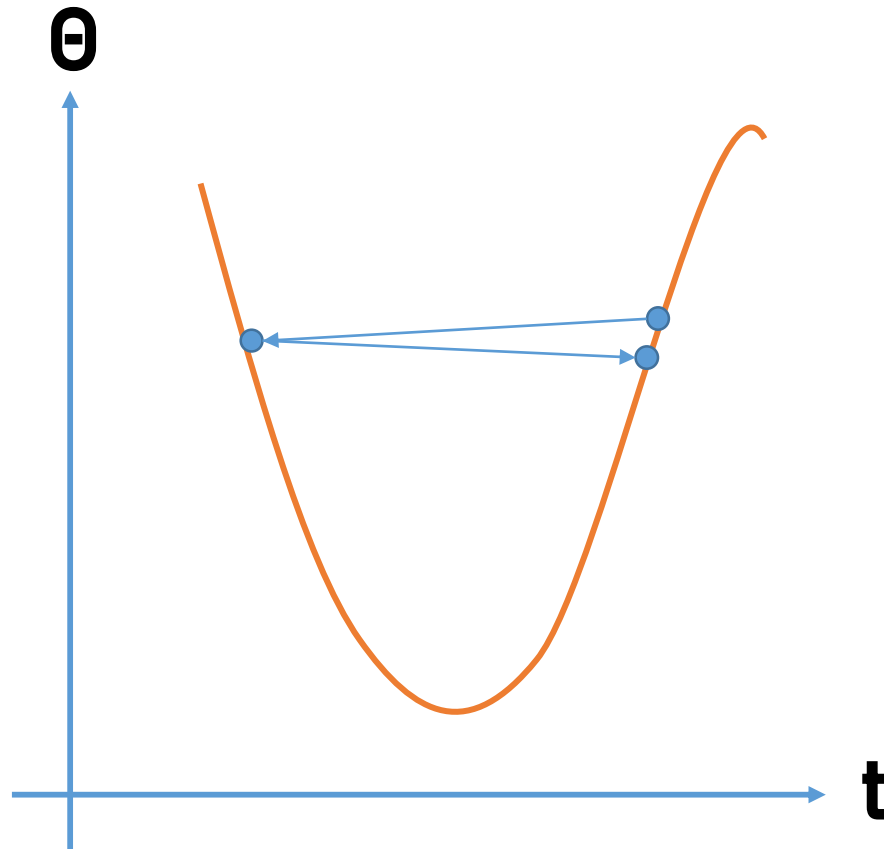
Very Small value of learning rate



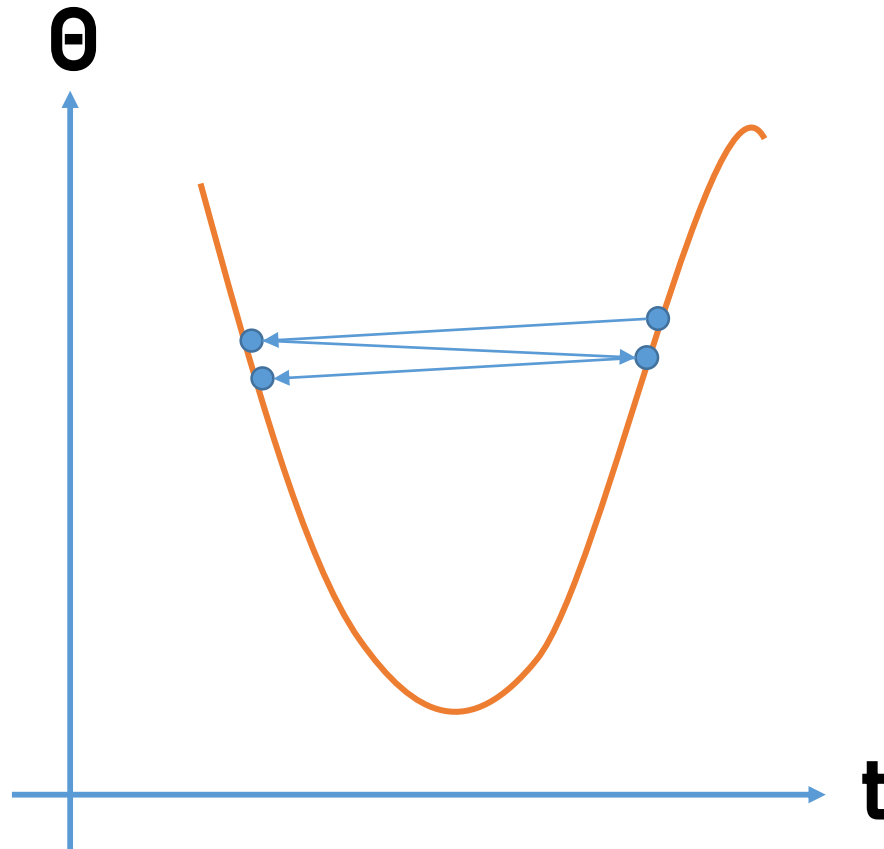
Very Small value of learning rate



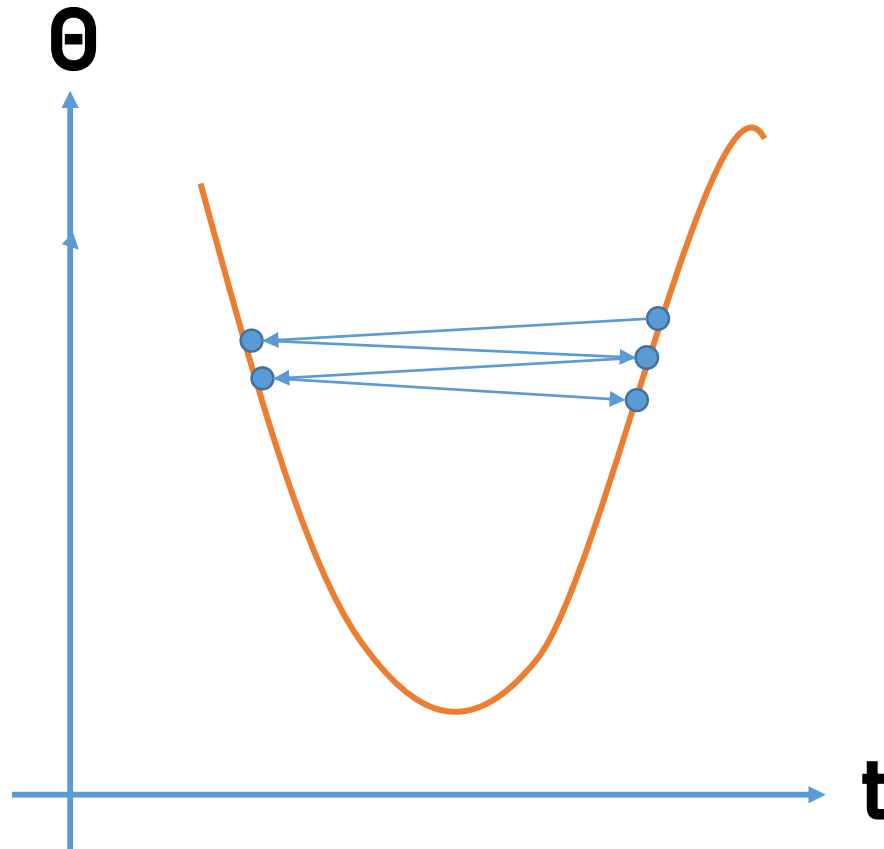
Very Small value of learning rate



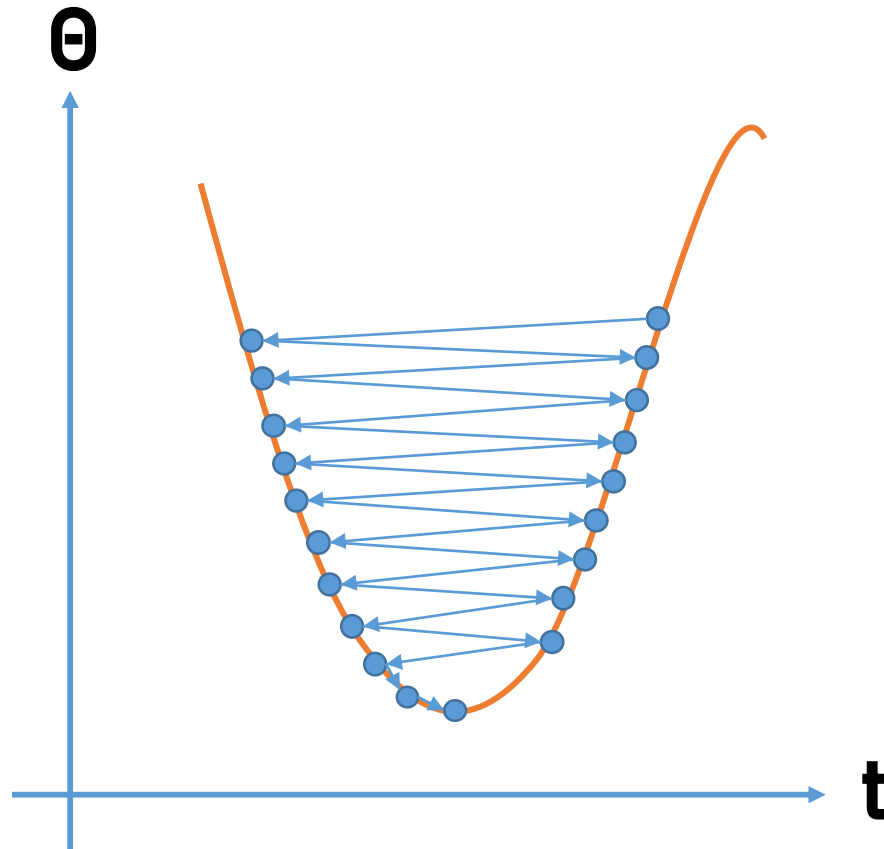
Very Small value of learning rate



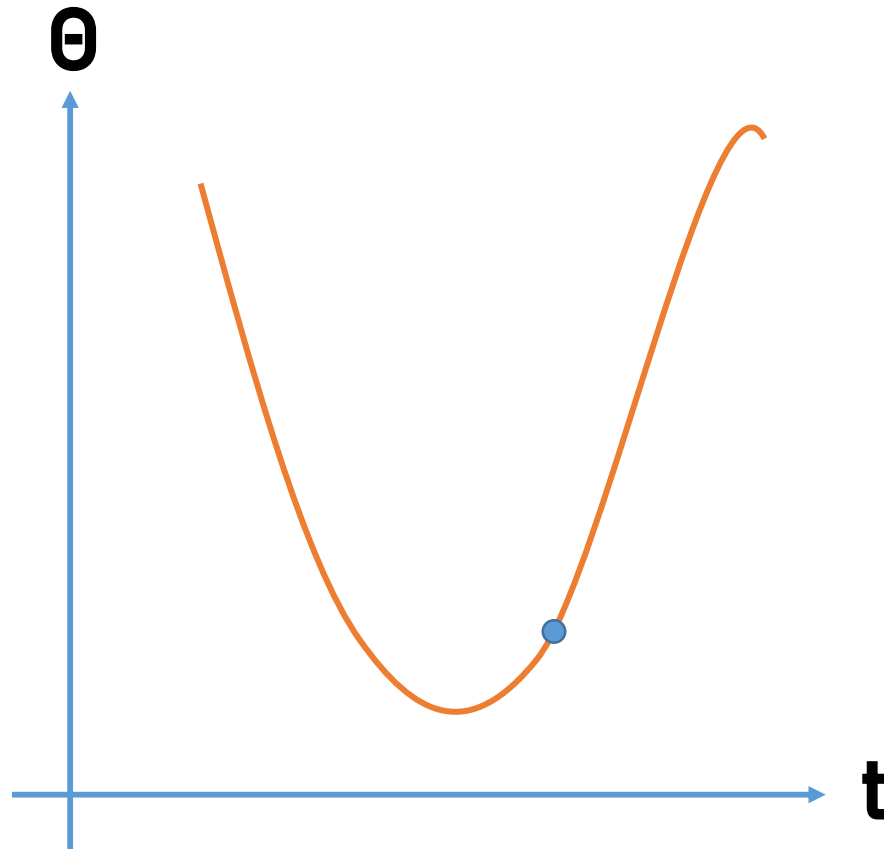
Very Small value of learning rate



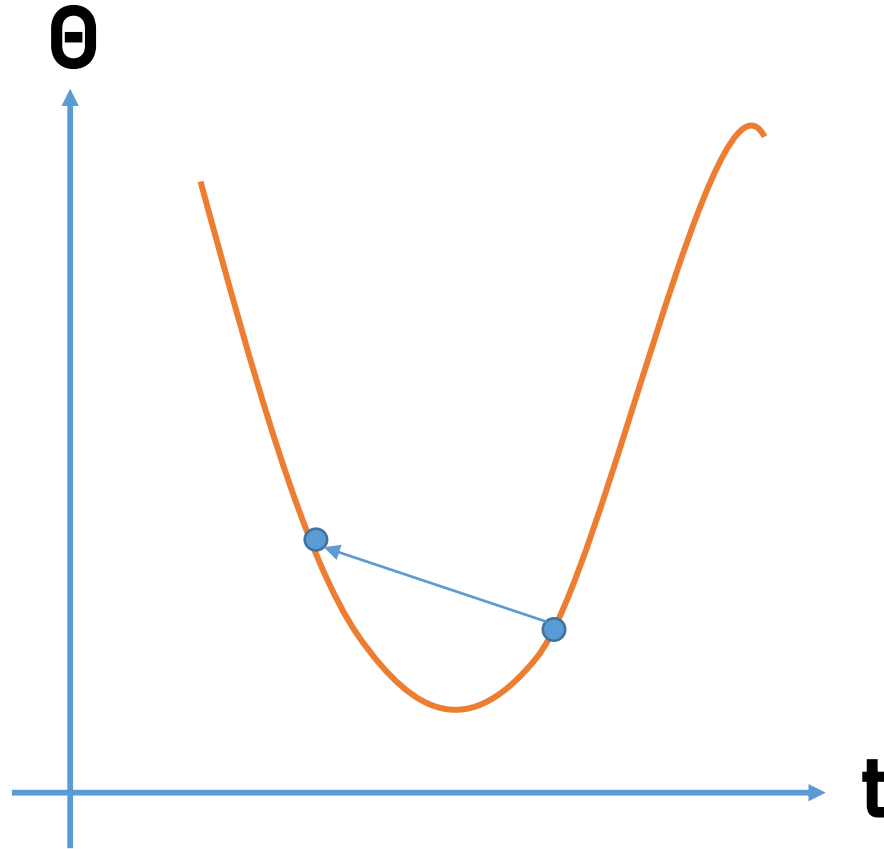
Very Small value of learning rate



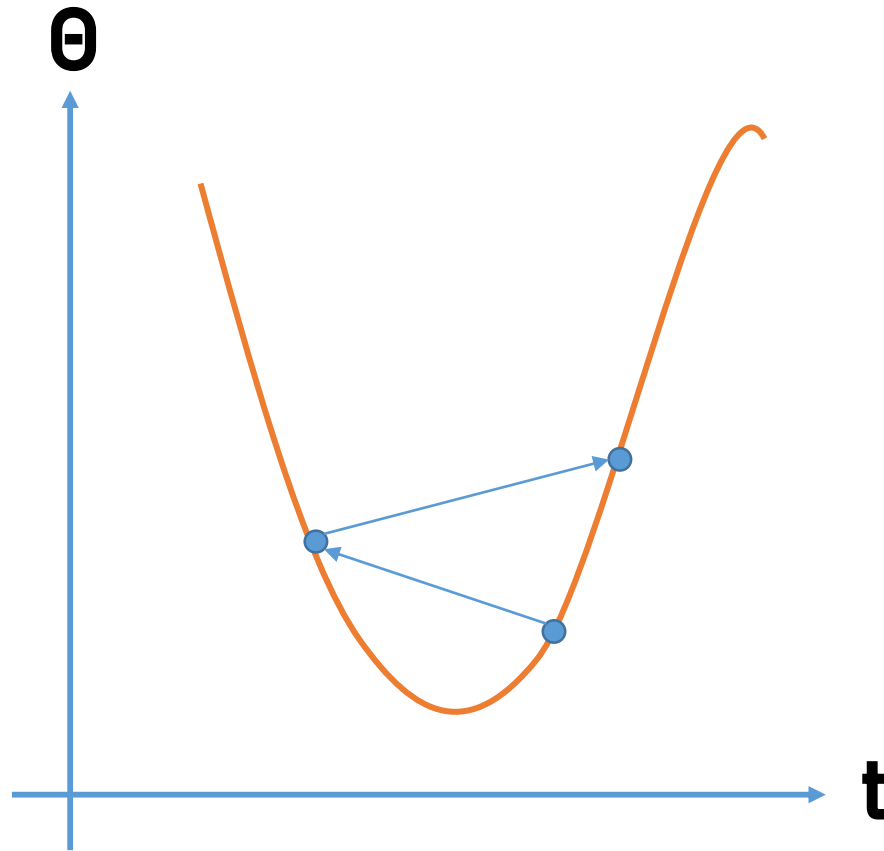
Very Large value of learning rate



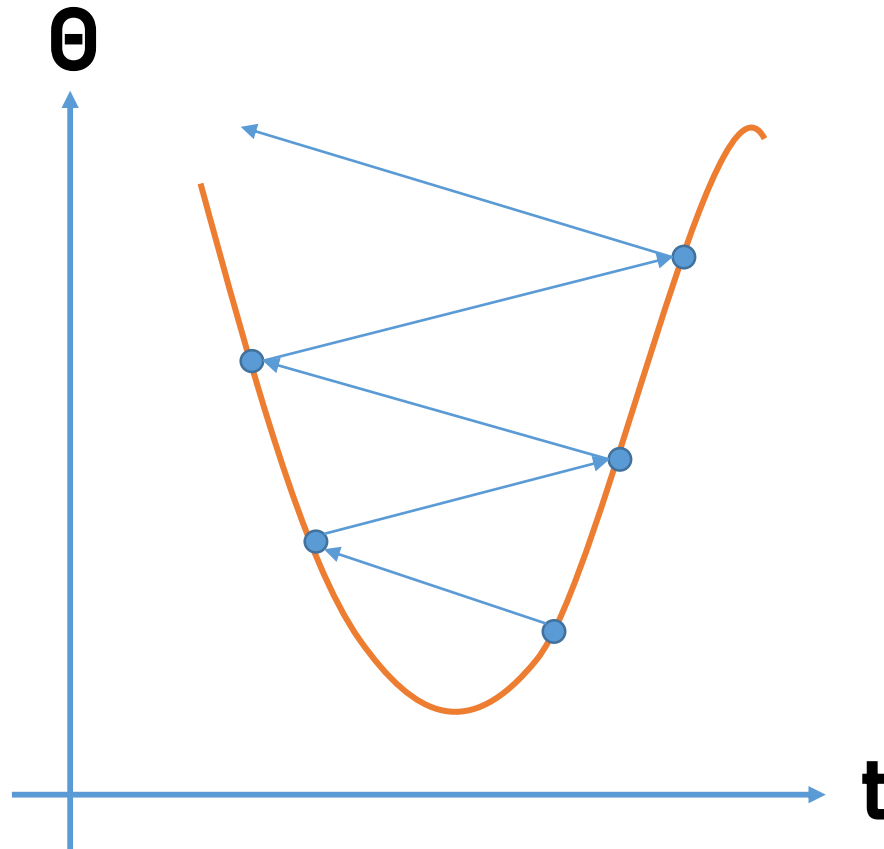
Very Large value of learning rate



Very Large value of learning rate



Very Large value of learning rate



Small but just right value of learning rate

