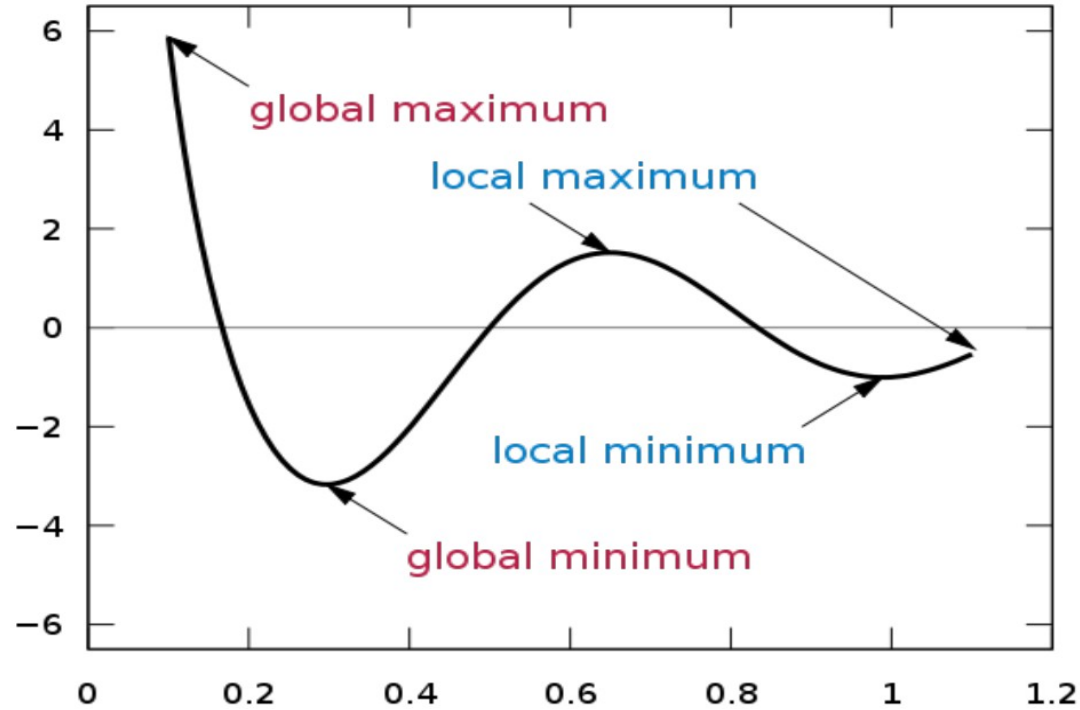
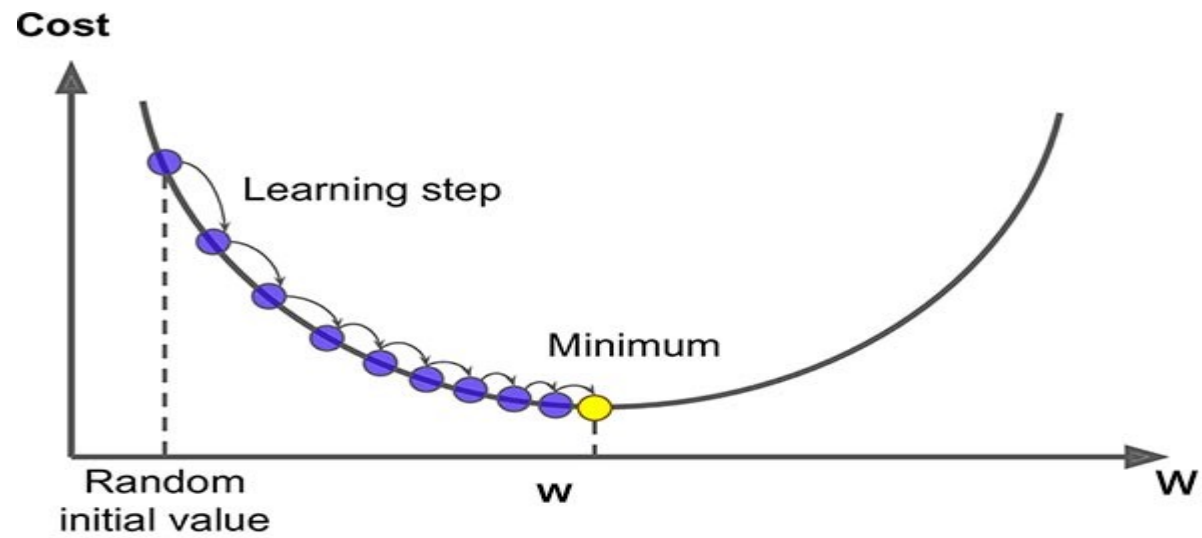
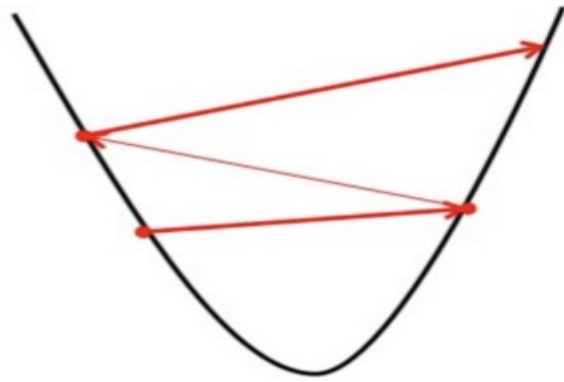


Function Example

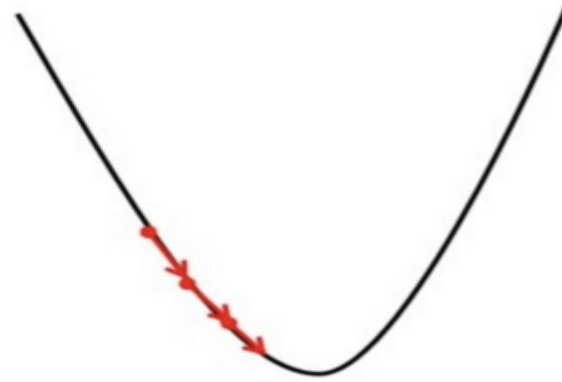


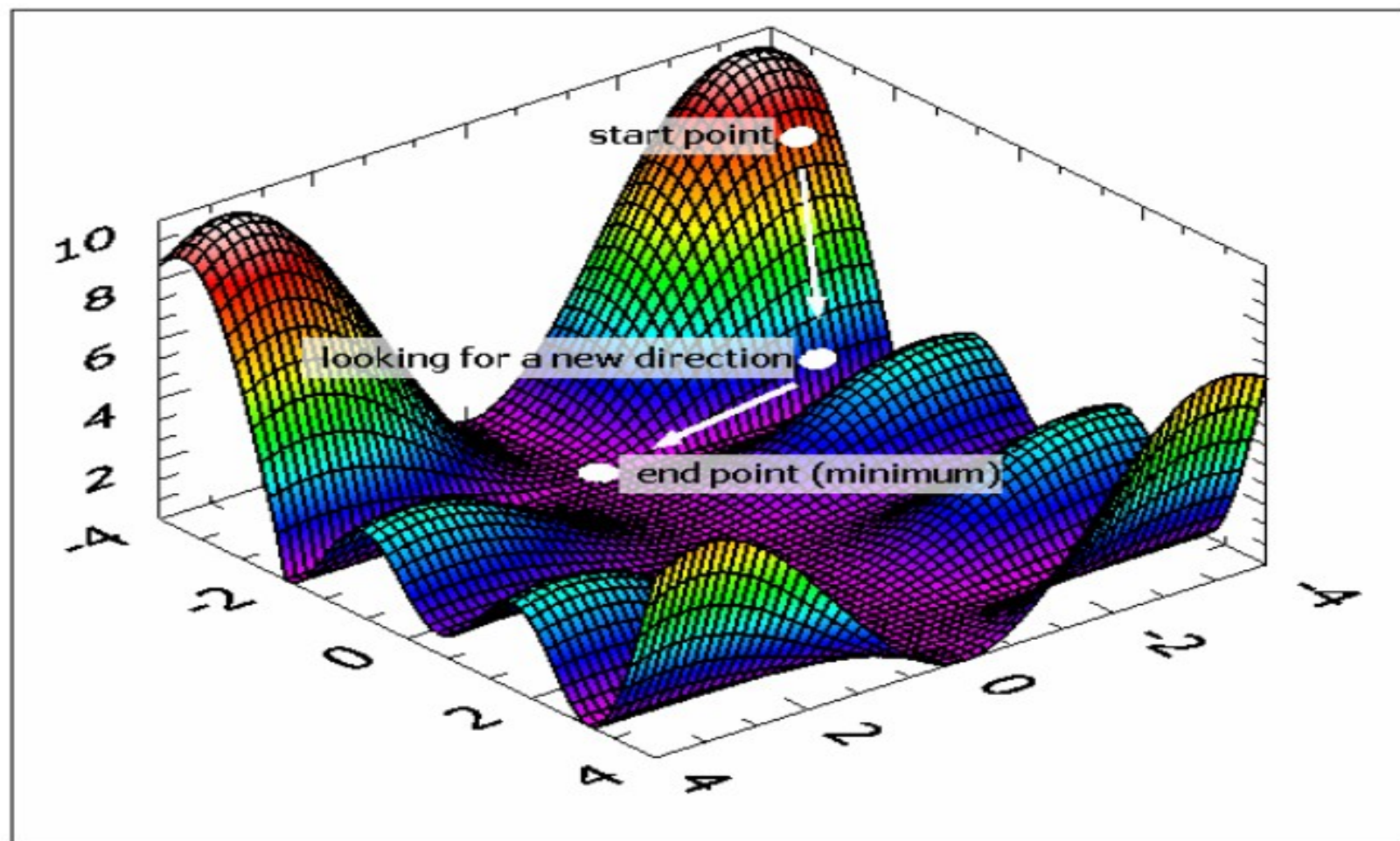


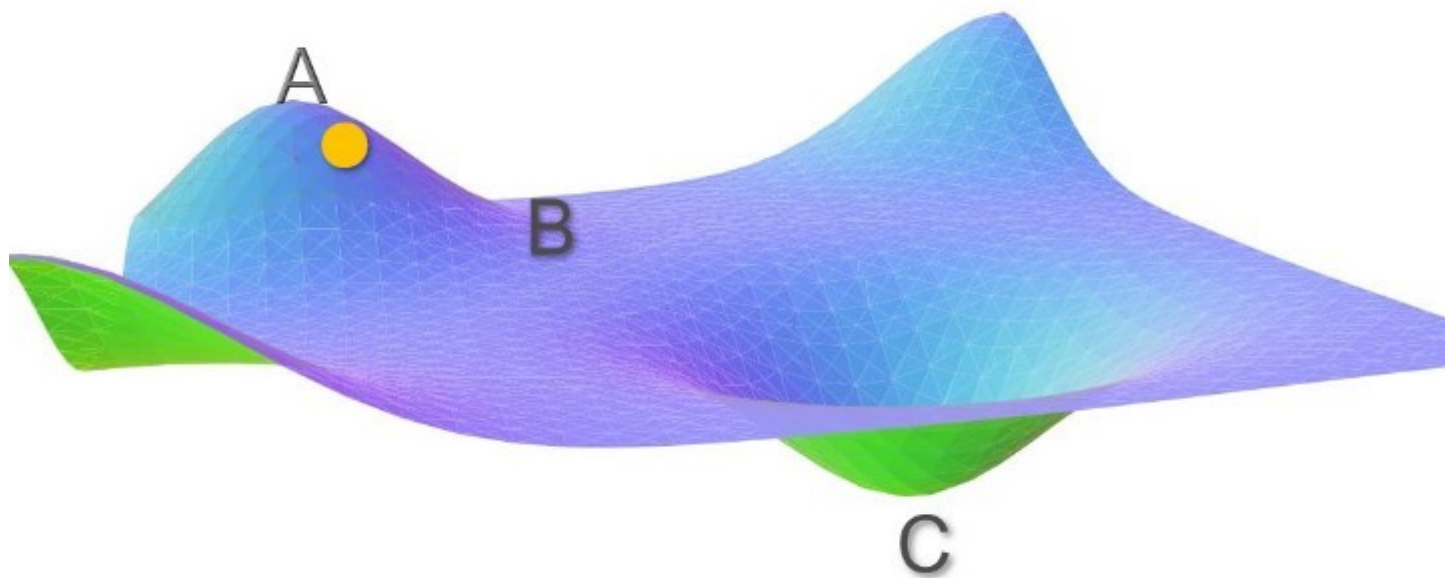
Big learning rate



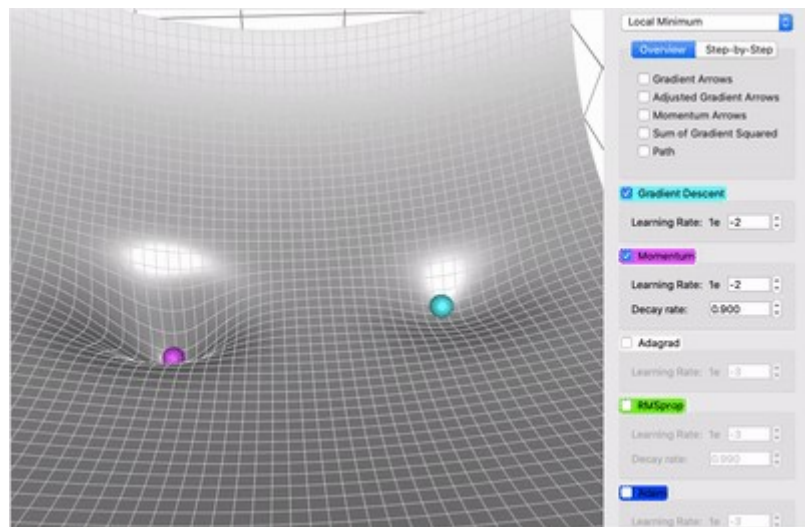
Small learning rate







Gradient descent with momentum

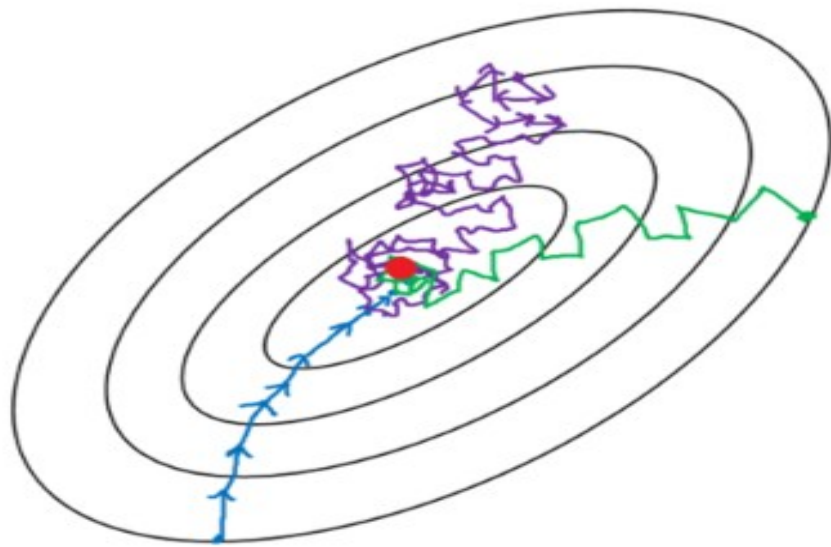


Batch Gradient Descent: Batch Gradient Descent uses the whole batch of training data at every step

Stochastic Gradient Descent: Stochastic Gradient Descent just picks one instance from training set at every step and update gradient only based on that single record

Mini-Batch Gradient Descent: combines the concept of Batch and Stochastic Gradient Descent. At each step, the algorithm compute gradient based on a subset of training set instead of full data set or only one record.





- Batch gradient descent
- Mini-batch gradient Descent
- Stochastic gradient descent