**梯度下降算法**

**(Gradient Decent Algorithm)**

# 梯度下降算法

## 批量梯度下降法（Batch Gradient Descent，BGD）



在更新每一参数时都使用所有的样本来进行更新

优点：全局最优解；易于并行实现；

缺点：当样本数目很多时，训练过程会很慢。

## 随机梯度下降法（Stochastic Gradient Descent，SGD）



在更新每一参数时都使用一个样本来进行更新

优点：训练速度快；

缺点：准确度下降，并不是全局最优；不易于并行实现。

## 小批量梯度下降法（Mini-Batch Gradient Descent，MBGD）



在更新每一参数时都使用一部分样本来进行更新

适用情况

如果样本量比较小，采用批量梯度下降算法。如果样本太大，或者在线算法，使用随机梯度下降算法。在实际的一般情况下，采用小批量梯度下降算法。

# 梯度下降优化算法

## 动量梯度下降（Gradient Descent with Momentum）





## Nesterov accelerated gradient





## Adagrad





## Adadelta（Adaptive Learning Rate）

## RMSprop（Root Mean Square Prop）

## Adam（Adaptive Moment Estimation）

### tf.train.GradientDescentOptimizer()

### tf.train.MomentumOptimizer()

### tf.train.AdadeltaOptimizer()

### tf.train.AdagradOptimizer()

### tf.train.AdagradDAOptimizer()

### tf.train.AdamOptimizer()

### tf.train.SyncReplicasOptimizer

### tf.train.RMSpropOptimizer()

### tf.train.FtrlOptimizer()

### tf.train.ProximalGradientDecsentOptimizer()

### tf.train.ProximalAdagradOptimizer()