# 深度学习中常用的优化算法：

<http://www.cnblogs.com/tornadomeet/archive/2013/05/02/3053916.html>

|  |  |  |
| --- | --- | --- |
|  | 特点 | 实验：在单核上比较SGD，LBFGS，CG三种算法的优化性能，当针对Autoencoder模型。结果如下：    结论：不同的优化算法有不同的优缺点，适合不同的场合（主要看目标函数），比如LBFGS算法在参数的维度比较低（一般指小于10000维）时的效果要比SGD（随机梯度下降）和CG（共轭梯度下降）效果好，特别是带有convolution的模型。而针对高维的参数问题，CG的效果要比另2种好。也就是说一般情况下，SGD的效果要差一些，这种情况在使用GPU加速时情况一样，即在GPU上使用LBFGS和CG时，优化速度明显加快，而SGD算法优化速度提高很小。在单核处理器上，LBFGS的优势主要是利用参数之间的2阶近视特性来加速优化，而CG则得得益于参数之间的共轭信息，需要计算器Hessian矩阵。 |
| 随机梯度下降（SGD） | * 实现简单 * 训练数据多时优化速度特别快 * 需要人为调整很多参数（学习率、收敛准则） * 序列方法，不利于GPU并行或分布式处理 |
| LBFGS（受限的BFGS） | * 参数维数比较低（<10000）时，比其他两个好，特别是带有convolutional的模型 |
| 共轭梯度法（CG） | * 高维的参数问题，CG的效果要比另2种好 |