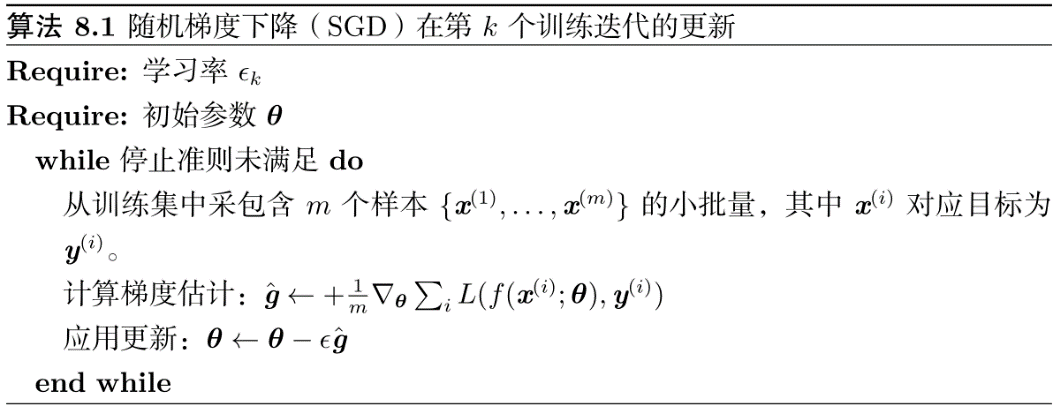
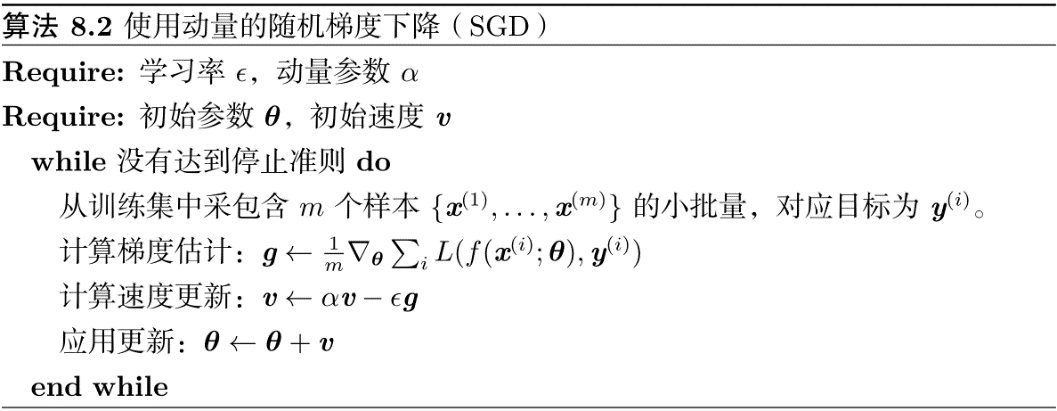
优化算法：

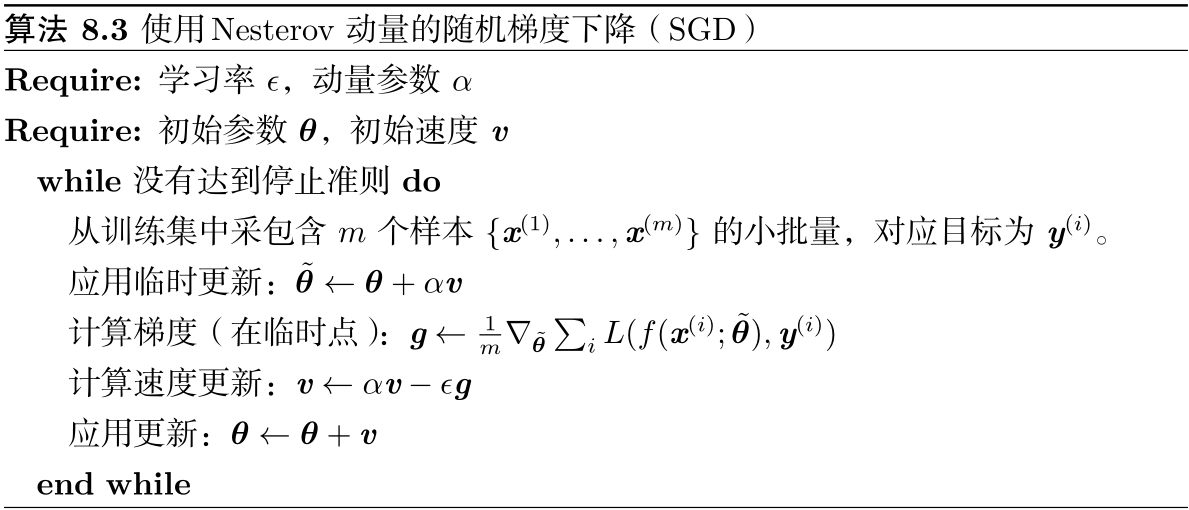
基本优化算法：SGD，动量，Nesterov动量

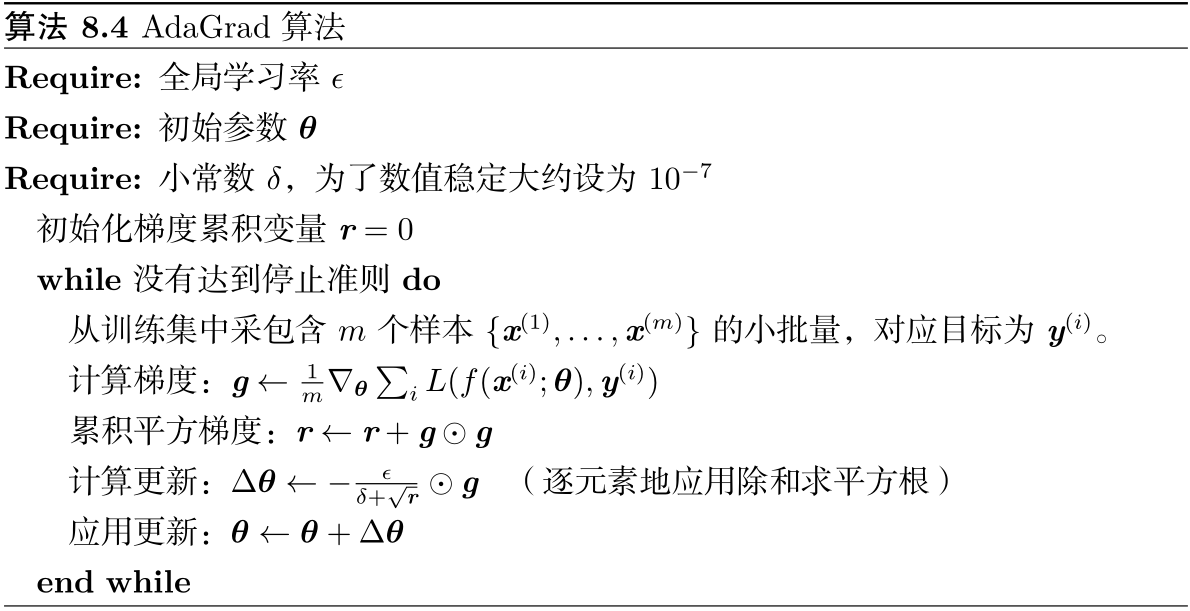
自适应学习率优化算法：AdaGrad，RMSProp，Adam

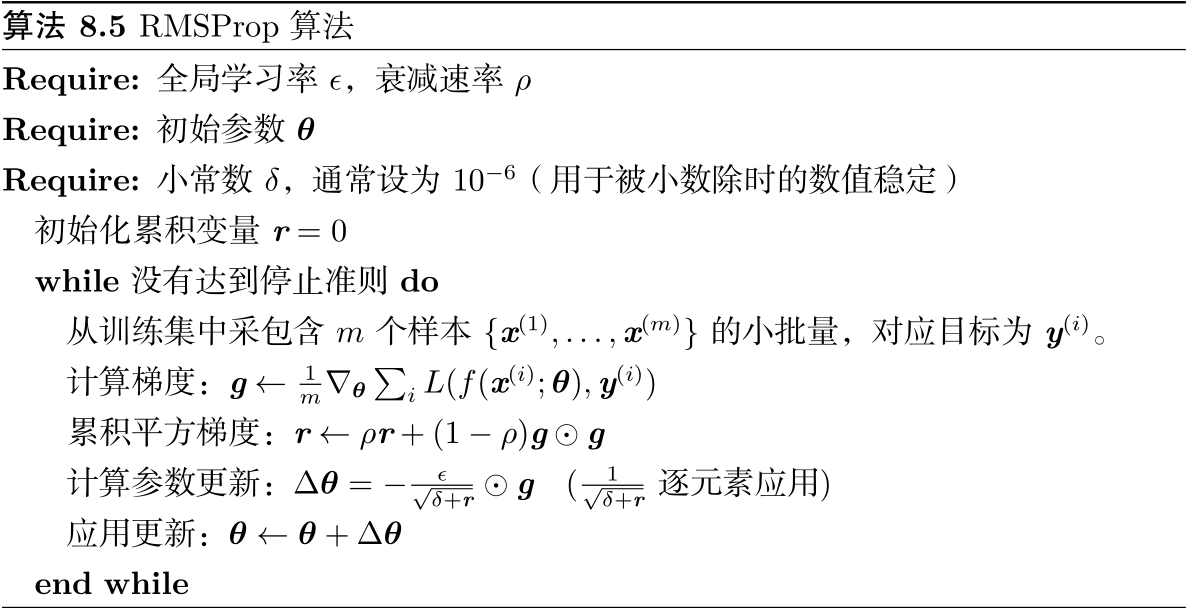
二阶近似优化方法：牛顿法，共轭梯度，BFGS

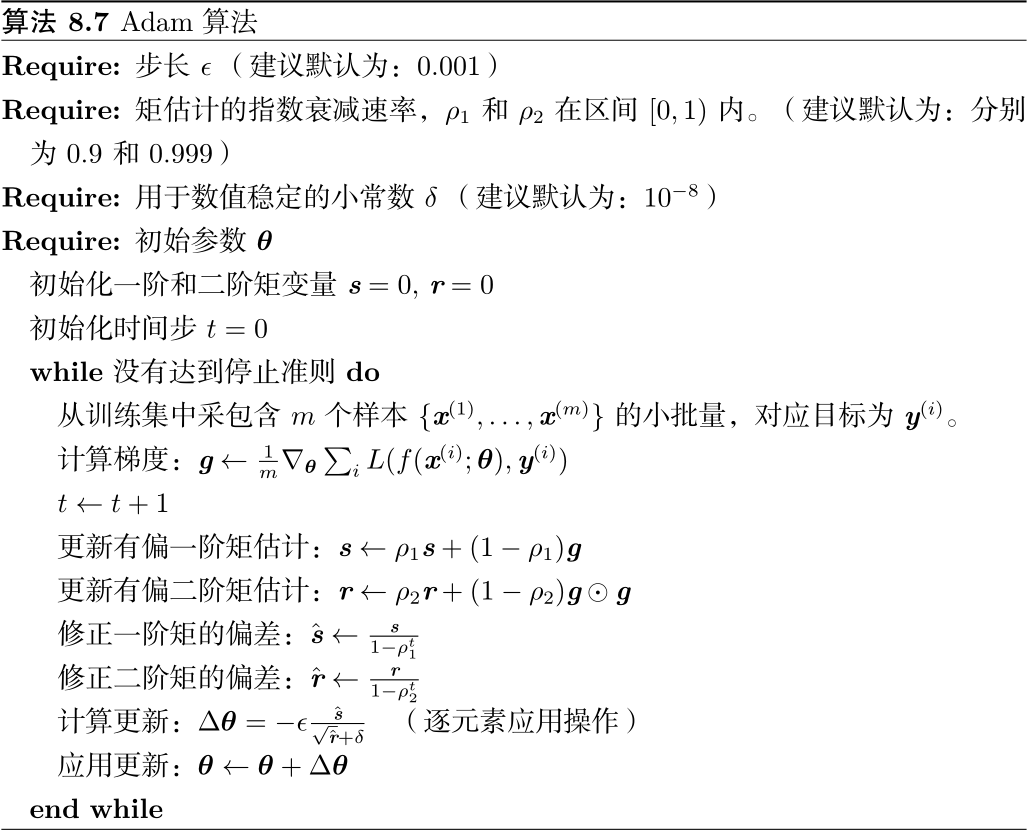












C:\Users\yanshuai\AppData\Local\Temp\1500295689(1).png 对theta求导得

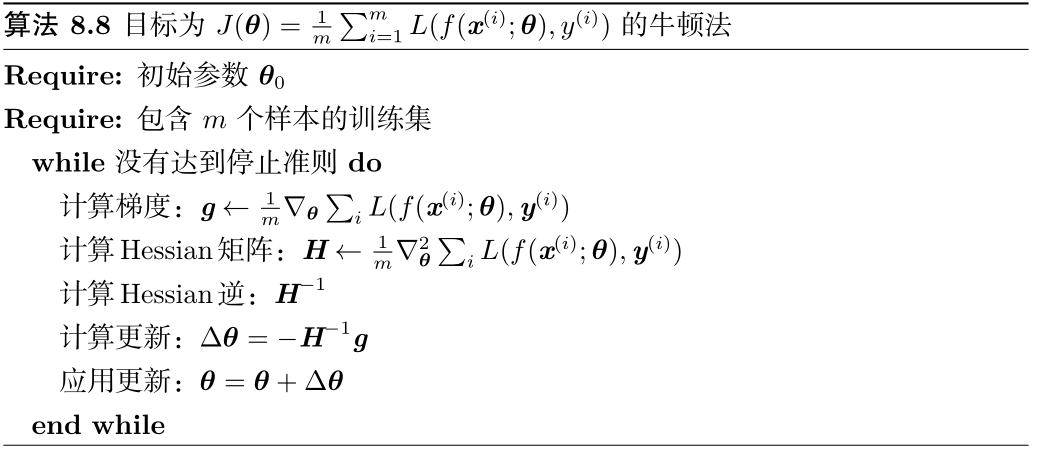
C:\Users\yanshuai\AppData\Local\Temp\1500295770(1).png

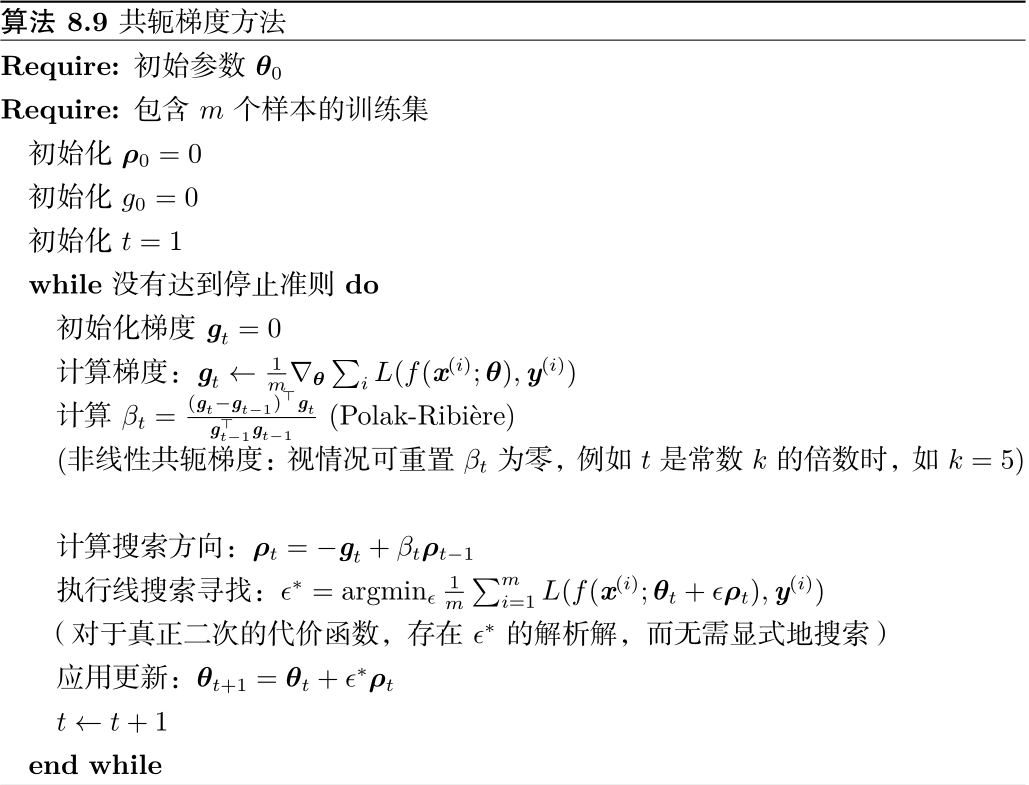
对于局部的二次函数（具有正定的 H），用 H −1 重新调整梯度，牛顿法会直接跳到极小值。对于非二次的表面，只要Hessian矩阵保持正定，牛顿法能够迭代地应用。

如果Hessian矩阵的特征值并不都是正的，这种情况可以通过正则化Hessian矩阵来避免。常用的正则化策略包括在Hessian矩阵对角线上增加常数 α。正则化更新变为



神经网络参数数目为 k，牛顿法需要计算 k × k 矩阵的逆，计算复杂度为 O(k 3 )。





拟牛顿法所采用的方法（BFGS是其中最突出的）是使用矩阵C:\Users\yanshuai\AppData\Local\Temp\1500284256(1).png近似逆，迭代地低秩更新精度以更好地近似 。