李航统计学习：

第五章讨论部分：

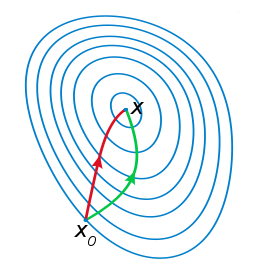
读书报告内容：

1. （必填）自己提出的问题的理解（罗列全部）：
2. 提出的问题1：

牛顿法和梯队下降法有什么区别和联系？

讨论后的理解：

关于牛顿法和梯度下降法的效率对比：  
a）从收敛速度上看 ，牛顿法是二阶收敛，梯度下降是一阶收敛，前者牛顿法收敛速度更快。但牛顿法仍然是局部算法，只是在局部上看的更细致，梯度法仅考虑方向，牛顿法不但考虑了方向还兼顾了步子的大小，其对步长的估计使用的是二阶逼近。

b）根据wiki上的解释，从几何上说，牛顿法就是用一个二次曲面去拟合你当前所处位置的局部曲面，而梯度下降法是用一个平面去拟合当前的局部曲面，通常情况下，二次曲面的拟合会比平面更好，所以牛顿法选择的下降路径会更符合真实的最优下降路径。  
  
注：红色的牛顿法的迭代路径，绿色的是梯度下降法的迭代路径。

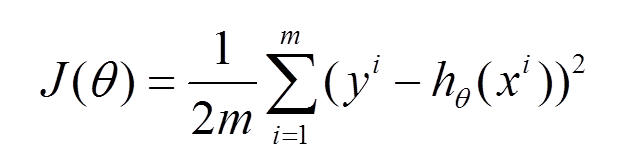
牛顿法的优缺点总结：  
优点：二阶收敛，收敛速度快；  
缺点：牛顿法是一种迭代算法，每一步都需要求解目标函数的Hessian矩阵的逆矩阵，计算比较复杂。

1. 提出的问题2：

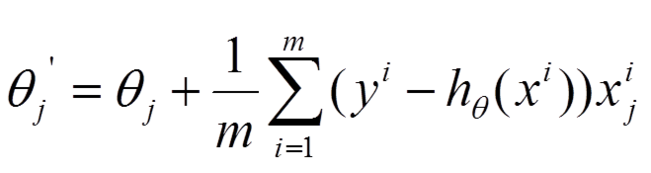
为什么凸函数时，梯度下降法是全局最优解？

讨论后的理解：

Linear Regression 的cost function如下：



是一个凸函数，也就是一个类似于开口向上的碗，它一定有最小值，并且只有一个局部极小值，那么你用梯度下降方法逐步更新theta的时候:

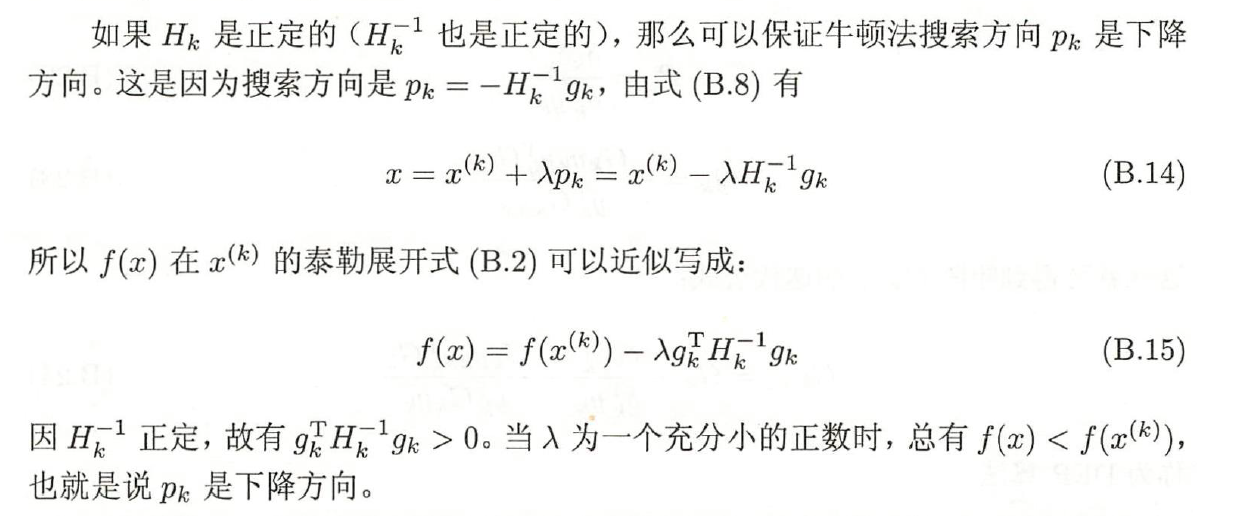


拟合函数最终一定会收敛到全局最优解。

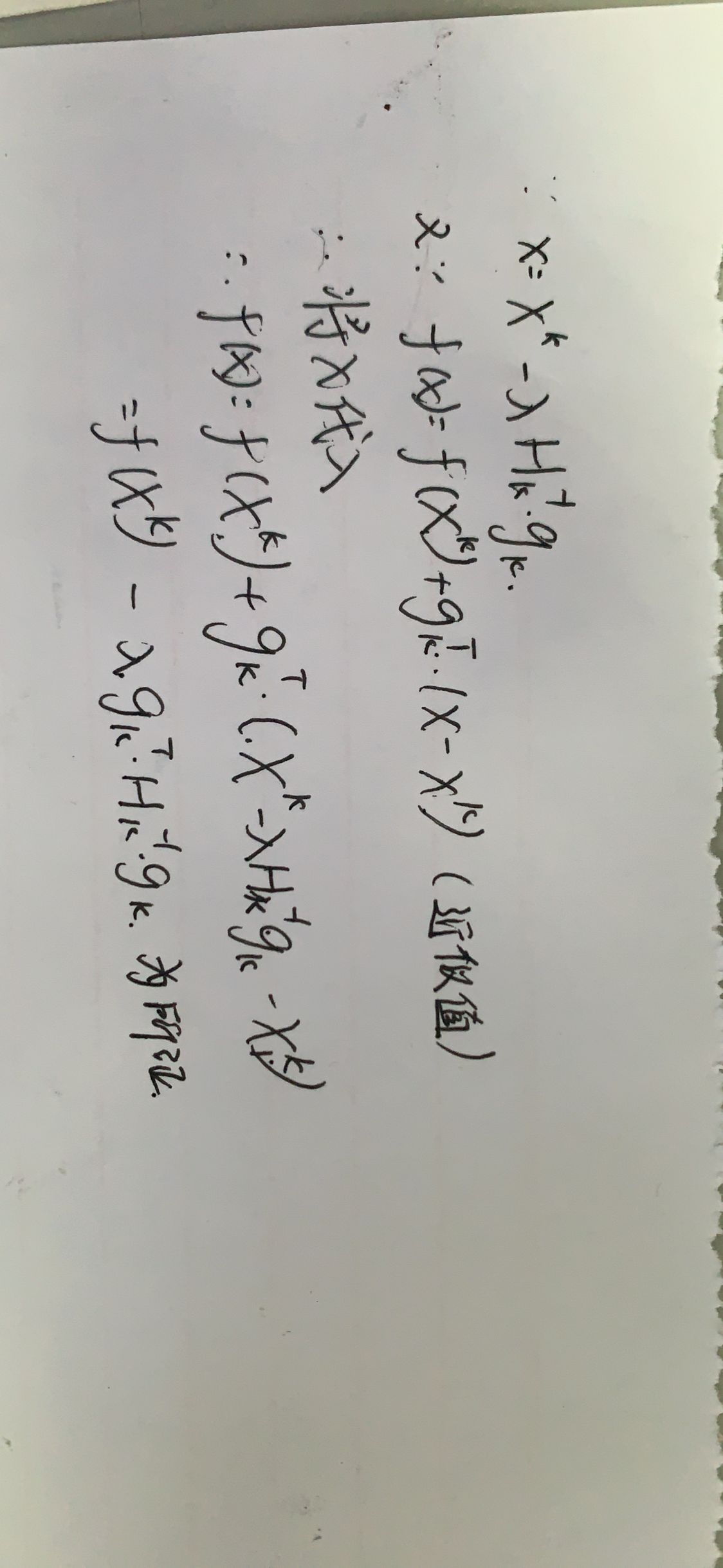
二、（必填）别人提出的问题的理解（选择几个问题罗列，并给出理解）：

问题3：如何从B.14推到B.15的？

自己的理解：



手写证明如下：



问题4：为什么“特别是当H是正定矩阵时，函数f的极 值为极小值”

自己的理解：

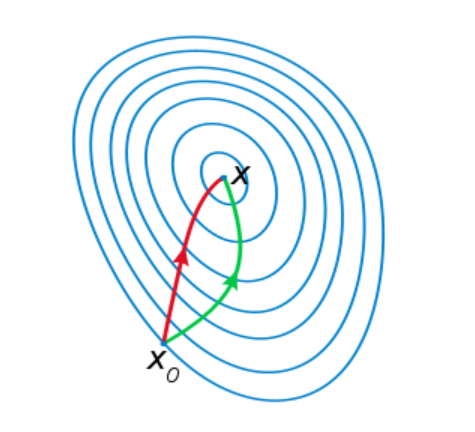
可以参考一维函数，函数取得极值的条件是一阶导等于零，二阶导不等于零。而矩阵求导X=A\*B，一阶导数为矩阵转置，二阶导数再次转置为本身，那么当H（二阶导）是正定矩阵，行列式大于0时，取得极小值。

问题5：从数学上来说，为什么牛顿法和拟牛顿法能够更快的找到极小值

自己的理解：

如下图所示，红色表示牛顿法，绿色表示梯度下降法。因为梯度下降法每次只从当前所处位置选一个坡度最大的方向走一步，牛顿法在选择方向时，不仅会考虑坡度是否够大，还会考虑走了一步之后，坡度是否会变得更大。所以，可以说牛顿法比梯度下降法看得更远一点，能更快地走到最底部。

从几何上说，牛顿法就是用一个二次曲面去拟合你当前所处位置的局部曲面，而梯度下降法是用一个平面去拟合当前的局部曲面，通常情况下，二次曲面的拟合会比平面更好，所以牛顿法选择的下降路径会更符合真实的最优下降路径。



1. （必填）读书计划
2. 本周完成的内容章节：

（1）附录ABC部分完成并且梳理知识点，寻找问题，自己思考并且于小组会之前完成了自己的思考。

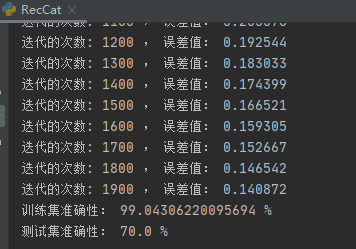
2、下周计划：第七章.

四、（选做）读书摘要及理解或伪代码的具体实现（读书摘要、伪代码的具体实现代码等可以写到这个部分）

搭建一个“识别猫”的简单的神经网络，实战运用了梯度下降模型：



运行结果：



跑一波的效果图是这样的，可以看出来成本在下降：

