李航统计学习：

第五章讨论部分：

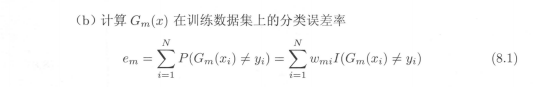
读书报告内容：

1. （必填）自己提出的问题的理解（罗列全部）：
2. 提出的问题1：

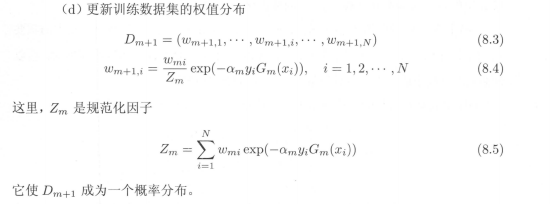
问题1：p158页中说由此可得Dm与Gm（x）的分类误差率的关系，没看懂是什么关系？

讨论后的理解：

Gm（x）的分类误差率为



Dm的概率分布为：

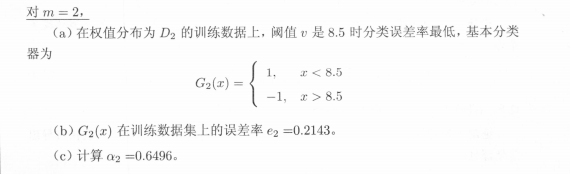


可以看出他们存在函数关系。

1. 提出的问题2：

p159页为什么阈值为8.5时分类误差率最低？

讨论后的理解：



对于权重D2的训练集上，阈值取8.5的时候，由于权值不同，加权之后的损失不同，所以选择阈值为8.5的时候，分类误差率最低，这里的分类误差率是加权后的分类误差率。

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）：

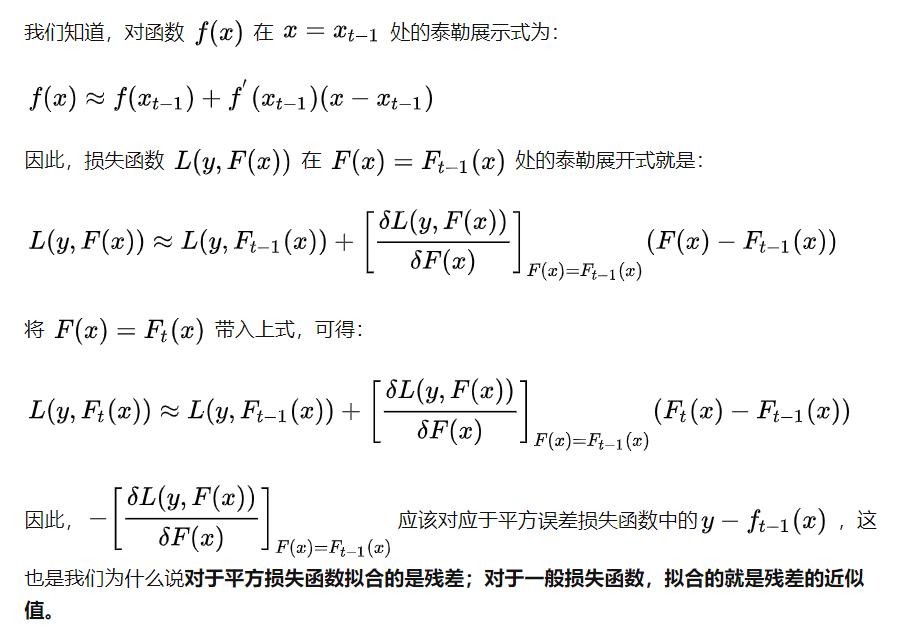
问题3：170页为什么提升树对指数损失函数和平方损失函数优化简单，损失函数是根据什么选择？

自己的理解：

指数损失函数感觉对应的是Adaboost算法而不是提升树算法，基函数为决策树桩的boost算法（就是提升树）使用了残差的概念来优化提升效果，而平方损失函数的负梯度正好是残差或残差的倍数。优化损失函数与正则项，其实代表的是对参数的极大似然或者极大后验估计，不同的损失函数和正则项，反映的我们对参数先验分布和似然函数的不同假设。

问题4：对于一般损失函数而言，为什么可以利用损失函数的负梯度在当前模型的值作为梯度提升算法中残差的近似值呢？

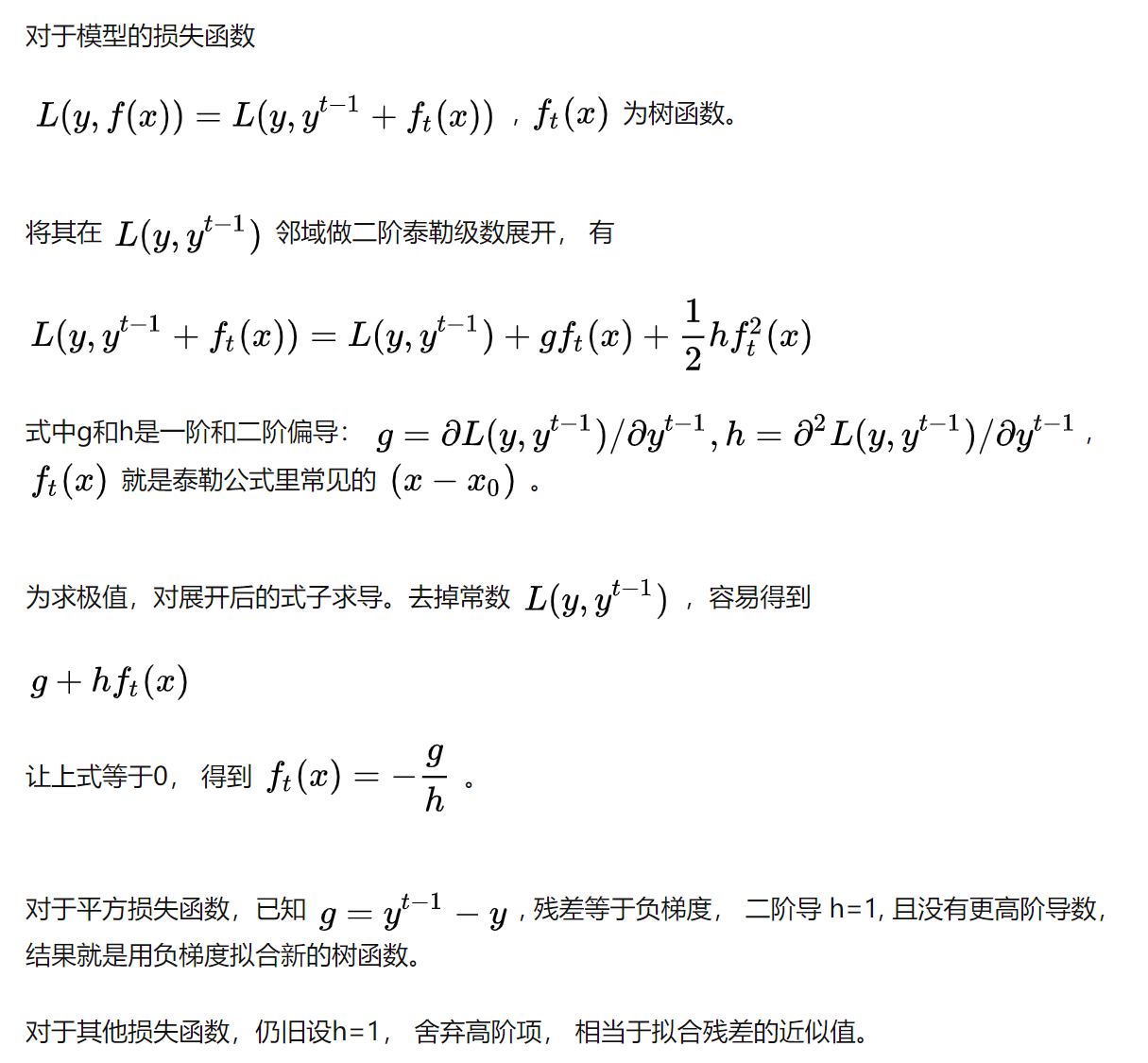
自己的理解：



即拉格朗日的取一介的近似值。

问题5：P171 为什么利用损失函数的负梯度作为残差的近似值？

自己的理解：



1. （必填）读书计划
2. 本周完成的内容章节：

（1）8.1到8.4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：第七章剩余.

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）

## **第一节 提升方法AdaBoost算法**

第一节首先介绍了AdaBoost的思想：是先得到一个弱分类器，然后在这个弱分类器的基础上将其提升到强分类器，具体方法是提升上一个分类器中被误分类的样本的权重，使得本次训练的分类器更加重视这些样本，最后的分类器是所有的分类器的线性组合。  
它的思想可以类比为秋招季参加了很多面试：第一次面试，面试官给你打了6分，并给你建议说你知识广度够，深度不够；第二次面试前你着重对知识的深度这方面进行了准备，第二次面试面试官给你打了6.5分，并给你建议说你的表达能力有限；第三次面试，你又对表达能力进行了准备，面试官给你打了7分。三次面试完后，你总结经验，着重第三次的经验，因为第三次面试官给你打的分最高。  
每一次面试，面试官给你建议就好像是算法中在该轮没有被正确分类的样本点，你开始重视这些建议就好比给这些误分类的点增加了权值，面试官给你打的分就好比是算法中该次弱分类器的权值，得分高的面试经验比得分低的面试经验更需要你重视。AdaBoost算法的思想就是这样，前一次没有正确分类的样本点在后一次会被更加重视，前一次的分类器训练结果是会影响后一次分类器结果的。  
AdaBoost先对所有划分选择出一个误分类最小的划分，得出一个分类器，分类器的权值

am=12log(1−emem)am=12log(1−emem)

，样本的权值也发生更新

wm+1,i=wm,iZme−amyiGm(xi)wm+1,i=wm,iZme−amyiGm(xi)

也就是说前一次的训练的结果会被这次的分类器产生影响。  
问：AdaBoost算法每一次训练的训练误差相对于上一轮是不是一定减少？  
答：首先区分训练误差和训练误差率，训练误差

e=1N∑i=1NI(G(xi)≠yi)e=1N∑i=1NI(G(xi)≠yi)

不同于分类误差率

em=∑i=1Nwm,iI(Gm(xi)≠yi)em=∑i=1Nwm,iI(Gm(xi)≠yi)

,被上一轮分类错误的样本增加了它的权值，从而使得下一轮分类器的训练重视这些被上一轮分类器误分的样本，权值更高的样本更有可能被正确分类，那么相应的分类误差率一定减小，训练误差不一定减少。  
延伸：分类误差率一定一直减少，那么样本权值

am=12log(1−emem)am=12log(1−emem)

一定一直增加，所以可以确定前一轮的分类器的权值一定小于后一轮的权值，

amam

一定小于

am+1am+1

。

## **第二节 AdaBoost算法的训练误差分析**

第二节介绍了AdaBoost的误差，误差要用到泰勒公式的知识，具体公式看附加知识中。我们知道训练误差率是不断减少的，那么训练误差也在学习过程中不断地减少（注意是学习过程中分类器的加法模型的训练误差一直减少，而后一个分类器的训练误差不一定比前一个分类误差的训练误差小）。训练误差是随着每加权一个弱分类器而减少的，那么我们要得出训练误差

e=1N∑i=1NI(G(xi)≠yi)e=1N∑i=1NI(G(xi)≠yi)

的性质，就需要使其小于一个值（在这里是

1N∑ie−yif(xi)1N∑ie−yif(xi)

）,推出后一个值的性质，就可以得出前者的性质，比如后者呈指数形式下降，前者小于后者，那么它也呈指数下降。同理，如果一个变量是一直增大的，那么要证明这个变量的性质，就要使其大于一个值，推出这个值得性质就可以得出变量的性质。  
问：在第二节中AdaBoost的训练误差界的证明中得出，AdaBoost的训练误差是以指数速率下降的，和AdaBoost的损失函数是指数函数有什么联系？可不可以得出训练误差是以指数速率下降，所以选择损失函数为指数函数？  
答：损失函数的意义是得出模型给出的结果和实际结果的偏离程度，AdaBoost的损失函数选择指数函数，是因为AdaBoost在第一节处理的是二分类问题，如果处理的是回归问题那么选择的是均分差损失函数，损失函数的选择只与处理的问题有关。所以训练误差的性质与损失函数的选择之间没有联系。

## **第三节 前向分布算法与AdaBoost**

第三节介绍了前向分布算法并从前向分布算法的角度来看AdaBoost。首先介绍了前向分布算法，前向分布算法的目标是训练一个加法模型

f(x)=∑m=1Mβmb(x;γm)f(x)=∑m=1Mβmb(x;γm)

是从前向后，每一步只学习一个基函数及其系数，而平常的分布算法是从m=1到M所有参数

βm,γmβm,γm

的优化问题简化为逐次求解各个

βm，γmβm，γm

的优化问题，一步一个脚印肯定比一步登天更容易实现，不是吗？  
AdaBoost也可以从前向分布算法的角度来看，不过要设定基函数为基本分类器，损失函数为指数损失函数。它的每一轮的分类器的训练是为了拟合残差，第一轮是为了拟合样本数据，后面都是为了拟合残差。由前向分布算法可以推导出第一节的分类器权值

amam

和第m+1轮的样本权值

am+1am+1

。