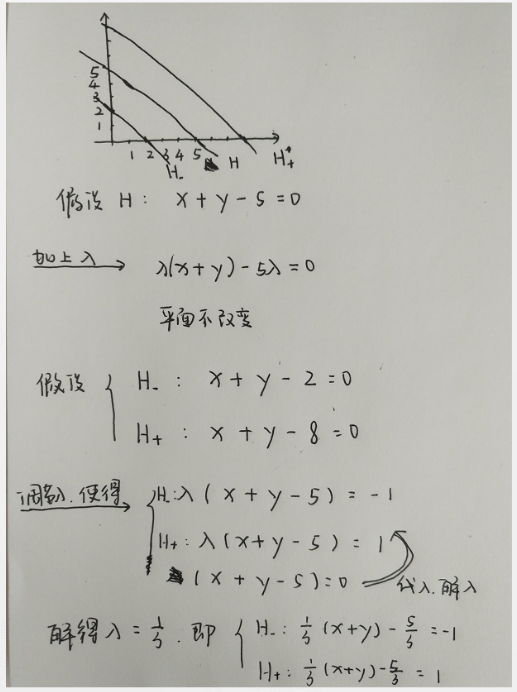
1. （必填）自己提出的问题的理解（罗列全部）：
2. 提出的问题1：P77拉普拉斯修正是否有可能对原先的预测结果造成影响

讨论后的理解：拉普拉斯修正能够有效避免因训练集样本不充分而导致估值概率为零的问题，而且因为n的取值是c的类别数，修正是比较平滑的。并且在训练集变大的时候，修正过程所引入的先验的影响也会变得可忽略，使得估值渐趋向于实际概率值。

1. （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）：
2. 问题2：83的第三行，通过加入参数，我们可以调节超平面，并且不改变函数，如何理解，参数为何一定是正实数？

自己的理解：加入参数是为了在平面不改变的情况下，将w 和b进行放缩，使得不改变超平面的正反，所以不使用负的参数

1. 问题3：p83式41下面说使用超平面上的任一点与边缘超平面的距离来作为d+，而不是x+到超平面的距离，为什么会这样计算？两种方式有什么本质上的区别？如果没有的话书上为什么会强调这一点？

自己的理解：书中P86中提到，在有约束优化问题中，我们使用所有的支持向量来计算b，取其平均作为最后的结果，而不是仅仅依靠一个支持向量，来减少数值误差。所以超平面上有可能没有支持向量，用d+就更为合理。

1. 问题4：软间隔svm中的损失函数为什么要选择hinge loss 这种形式，hinge loss有什么好处和局限性，有没有其他可比较的loss函数？

自己的理解：hinge loss有两个好处。一十实现了软间隔分类（这个Loss函数都可以做到），二是保持了支持向量机解的稀疏性，换用其他的Loss函数的话，SVM就不再是SVM了。

正是因为HingeLoss的零区域对应的正是非支持向量的普通样本，从而所有的普通样本都不参与最终超平面的决定，这才是支持向量机最大的优势所在，对训练样本数目的依赖大大减少，而且提高了训练效率。

1. 问题5：拉格朗日乘数法的现实意义

自己的理解：SVM问题是一个不等式约束条件下的优化问题。拉格朗日乘数法可以用来解有约束优化问题的最优解。

1. （必填）读书计划

1、本周完成的内容章节：3.6-3.8

2、下周计划：3.9-3.10

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）

1、读书摘要及理解（选做）

3.6朴素贝叶斯分类是一种十分简单的分类算法，叫它朴素贝叶斯分类是因为这种方法的思想真的很朴素。朴素贝叶斯的思想基础是这样的：对于给出的待分类项，求解在此项出现的条件下各个类别出现的概率，哪个最大，就认为此待分类项属于哪个类别。为避免因训练集样本不充分而导致概率估值为零，可以使用拉普拉斯修正。

3.7朴素贝叶斯文本分类把每篇文档看作一袋词，每个文档可以看成多项式分布生成。然后进行参数估计，这是基于经验的计数得到的。为了解决不常出现的单词的零概率估计，要使用拉普拉斯光滑。确定估计的参数后计算测试文档属于每个类别的概率。如果最后的分类器是将每个文档放到一个类别中，那么有最大后验概率的那个类别就是最后选择。



3.8支持向量机分类的方法就是希望能够找到数据集中离分隔超平面最近的点，确保他们离分隔面的距离尽可能的远。那个具有最大间隔的分隔超平面就是SVM要寻找的最优解。在SVM是线性可分的时候，由于目标函数是二次的，而约束在参数w和b上是线性的，因此这个问题是一个凸优化问题，可以通过标准的拉格朗日乘子方法求解。而非线性可分的SVM可以通过核函数，将数据映射到高维空间进行分类。简单的说，首先是使用核函数，使用一个非线性映射将数据变换到一个特征空间，然后在特征空间使用线性学习器分类。