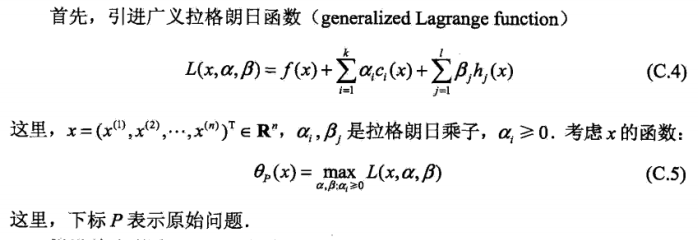
读书报告

71117203袁佳怡

1. （必填）自己提出的问题的理解（罗列全部）：
2. 提出的问题1：为什么使用核函数，核函数的本质是什么？

讨论后的理解：因为在实际中，我们会经常遇到线性不可分的样例，我们常用的做法是把样例映射到高维空间中去，映射到高维空间后，相关特征便被分开了，就达到了分类的目的。但是如果遇到线性不可分的样例，一律映射到高维空间，那么这个维度大小可能会很高，此时就需要核函数，核函数的价值在于它虽然是将特征进行从低维到高维的转化，但核函数事先在低维上进行计算，而将实质上的分类效果表现在了高维上，避免了直接在高维空间中的复杂计算。

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：



原问题是min，θp(x)是max，为什么能表示原问题？

自己的理解：这里的θp(x)表示的是f(x)，需要让f(x)得到最大，就是为了得到最大间隔的分类器。

1. 问题3：支持向量的作用，是不是把正例点和负例点尽量分割的开一些？

自己的理解：在支持向量机中，距离超平面最近的且满足一定条件的几个训练样本被称为支持向量。就是离最优分类平面最近的离散点，也可以称为向量。超平面的参数完全由支持向量的点确定，超平面和任何其他的点无关，如果改变其他点的位置，只要其他点不落入虚线上或者虚线内，那么超平面的参数都不会改变。

1. （必填）读书计划

1、本周完成的内容章节：如2.1-2.6

2、下周计划：第三章

四、读书笔记

1、支持向量机是一种二类分类模型。支持向量机的学习策略就是间隔最大化，可以形式化为一个求解凸二次规划的问题，也等价于正则化的合页损失函数的最小化问题。支持向量机的学习算法是求解凸二次规划的最优化算法。与感知机相比，支持向量机的要求更高，不仅仅要求正确分类，还要求正确性尽可能高，正确性表现为样本点到超平面的距离（间隔）。具体策略与模型的选取有关，又分为硬间隔最大化和软间隔最大化。

2、线性可分支持向量机。需要找到能将两类数据分隔开的直线。离超平面距离越远的点被正确分类的距离越高，因为和分界线的安全距离更大。这个安全距离也叫做函数间隔。对训练数据集而言，训练集的函数间隔是所有样本点的函数间隔的最小值。将函数间隔归一化，就是几何间隔。线性可分支持向量机的策略就是基于几何间隔的，使几何间隔最大化，就是确保分隔的可信度最高。是训练集的集合间隔最大并且同时保证训练集中每一个样本的几何间隔都大于训练集的集合间隔。经过等价转换，转化为一个凸二次规划问题（目标函数和约束函数都是连续可微的凸函数），这样在超平面两侧各有一个新的超平面wx+b=1和wx+b=-1，落在这两个超平面上的样本点决定了分离超平面，被叫做支持向量。应用拉格朗日的对偶性，构造并求解另一个约束最优化问题，然后再求解w和b。

3、线性向量机 在这种情况下，去除少部分特异点，其他大部分样本点是线性可分的，通过降低约束条件来实现去除特异点，具体做法是引入一个松弛变量，但同时对目标函数进行修改，引入一个待见，保证在间隔尽量大的同时误分类点尽量少。由此得到线性向量机的学习算法。再线性支持向量机中，支持向量不再只分布在约束条件边界上，而是还分布在间隔边界和分离超平面之间。并且书中还提到了近似线性可分数据的另一种解决思路即利用合页损失函数，思想是在同时满足分类正确和间隔大于约束条件时损失才是0，否则损失是1-y(wx+b).

4、非线性支持向量机，不是所有的数据都是线性可分的，有的数据线性不可分，却可以利用非线性模型如椭圆分离开来，这是我们希望将非线性问题变换为线性问题。非线性支持向量机的重点主要在非线性映射的问题。支持向量机是在特征空间上进行分类的。具体的映射关系可以通过核函数来实现。核函数是输入空间的分类曲线，在选取核函数解决实际问题时，通常采用的方法有：一是利用专家的先验知识预先选定核函数；二是采用交叉验证，即在进行核函数选取时，分别试用不同的核函数，归纳误差最小的核函数就是最好的核函数。

讲座报告

因为自己课程的时间冲突，我只参加了第二次的爱思维尔的讲座，接近六个小时的讲座也是内容丰富让人受益匪浅。整个疫情下，我们的国家在各个方面都面临着各种挑战，我们更是集中各方面的力量来走出这场困境，这是生命的诘问，也是历史的考题。

在这次的讲座中，几位大咖为我们讲述了疫情之下的一些研究工作和它们的使用价值，整个报告围绕着疫情挑战下的大数据和人工智能研究展开，不管是达摩院的医疗健康AI的设计思想、还是新型冠状病毒肺炎CT辅助诊断系统方法的落地使用，又或者说是利用自然语言处理技术的智能问答及自动翻译等各种举出的案例的使用无不是技术与知识的力量改变着传统的医疗工作体系。在我看来，智能AI具有强大的发展前景，将能够被广泛地使用，在病情监控、病理分析等各个方面大显神通。

在整个讲座中令我最为印象深刻的是清华计算机系唐杰老师团队做的基于AI驱动的新冠肺炎知识服务的知识疫图，在这个系统中，他们团队致力于汇聚冠状病毒的各种数据源，实现基于大数据的智能预测，并且构建起冠状病毒的知识图谱。他们以官方公布数据为基础，建立了传统传染动力传播SEIR模型，引入各种影响因子，然后构建了一个疫情趋势预测模型，在疫情防控战中积极科研，展现了作为一线科研人员的力量。

其实技术的力量在于切实解决了问题，这些教授们所介绍的应用就是因为这一点方显非凡价值。这些领先的技术让我非常震撼，他们的那种科研精神也令我非常崇敬。