山东大学软件学院2018机器学习期末考试（回忆版）

总分一百分，开卷。

一、名词解释（5\*4）

1、机器学习

2、主动学习

3、ID3

4、神经网络

二、简答题（10\*3）

1、简述Parzen窗方法的原理和过程。证明为什么可以用高斯概率密度函数作为窗函数。

2、简要说明梯度下降法和牛顿法的基本思想和区别。解释为什么梯度下降法能够保证一定是下降的？

3、什么是过拟合？为什么会出现过拟合？如何解决过拟合问题？

三、综合题（20+30）

1、证明AdaBoost算法的合理性。

2、在SVM支持向量机中，简述为什么要使margin最大化？求下列函数的对偶函数（PPT上原公式）。对于非线性问题该如何解决又什么方法吗？

————————————————

版权声明：本文为CSDN博主「ForestMonkey」的原创文章，遵循CC 4.0 BY-SA版权协议，转载请附上原文出处链接及本声明。

原文链接：https://blog.csdn.net/qq\_37158580/article/details/85263438

山东大学软件学院2019机器学习期末试题整理（数媒）

试卷共六个大题，基本都在老师考前画的重点内，不过还是有些出乎意料的…可能因为往年开卷的原因，感觉复习资料特别少，简单整理一下供大家参考。

自己整理的复习资料，欢迎下载：添加链接描述

一、 概念题

1、机器学习过程，每个环节主要操作

2、集成学习的概念，BOOSTING和BAGGING 思想

3、交叉验证法

4、特征提取和特征选择的区别

5、隐马尔科夫模型定义形式，以及解决的三个问题？

6、贝叶斯决策过程

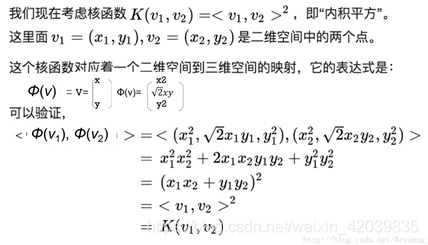
二、 svm

1、svm概念,其目的,什么是最优化分类面

2、最优拉格朗日形式

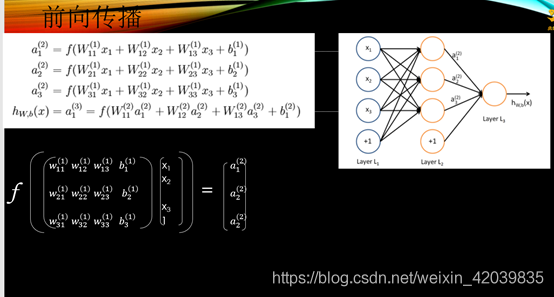
3、核函数为什么要引入函数

4、验证核函数，基本如下，只是x,y换成了x1,x2，映射函数是（x^2 x 1/2）T



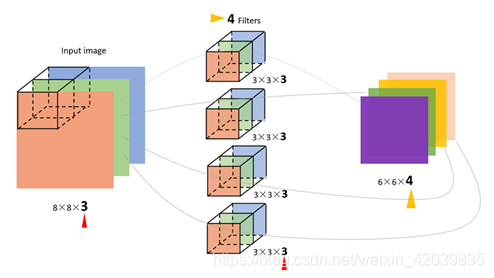
三、 神经网络、

1. 描述bp算法
2. 前向传播表达式a1
3. 输出层输出形式  
   （嗯，又一个原封不动的）



四、 深度学习

1. 卷积层作用
2. 激活函数以及形式
3. 输出  
   给出下图左1，左二，问你卷积后输出几个，大小为多少，（输出即右一图，这页ppt我根本就没瞅啊，现在就是后悔！！！）



五、 决策树

1.决策树算法思想是什么，两个分类

2.给出一个表格

写出预处理数据集

决策树缺失项 以及依据

决策树分类规则

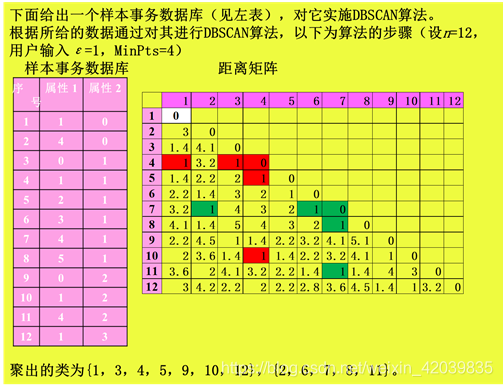
给数据判断结果

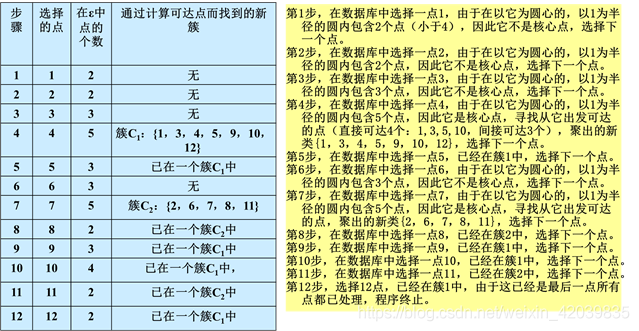
给出的表格决策树基本如下：





六、 DBSCAN算法应用计算（给距离矩阵求聚类结果）





————————————————

版权声明：本文为CSDN博主「壹零壹零壹零壹」的原创文章，遵循CC 4.0 BY-SA版权协议，转载请附上原文出处链接及本声明。

原文链接：https://blog.csdn.net/weixin\_42039835/article/details/103478413

2021山东大学机器学习期末考试(回忆版)

这是限选课的机器学习，期末只占40分，开卷，强烈建议带一本书(模式分类或西瓜书都行，我个人考试时基本只用到西瓜书)

名词解释

knn

独立同分布

间隔(margin)

激活函数

简答

什么是outlier，为什么k均值对outlier敏感

一道朴素贝叶斯，为应该分到哪类

一道决策树，根据信息增益判断选哪个属性

什么是线性可分，svm怎么解决线性不可分问题，简述思路

综合

什么是训练误差，什么是泛化误差

画图说明误差，指出过拟合和欠拟合区域

梯度下降为什么选择负梯度优化目标函数

怎么选择学习率，过大过小的影响?

综合

验证集的作用，怎么使模型学习到全部数据

adaboost，权重α的推导

————————————————

版权声明：本文为CSDN博主「Ryan-J」的原创文章，遵循CC 4.0 BY-SA版权协议，转载请附上原文出处链接及本声明。

原文链接：https://blog.csdn.net/m0\_46306466/article/details/122370851

2020-2021年度山东大学软件学院AI班机器学习考试（大三）

（与软工和数媒题不一样，但可以参考一下）

考试时间：2020年12月31日 14:00-16:00

考试形式：开卷，允许带计算器

先说一下感受，开卷考相比之前的“贝多芬（背多分）”好了很多，主要靠理解，这也正是我喜欢的形式。除了两个计算贝叶斯和决策树以及最后一个简答题之外，其余所有的题都可以在西瓜书和自己带的资料中找到答案。贝叶斯计算极简单，并且也是PPT最后几页中的homework，决策树计算被我们“3人机器学习小组”的大神直接压中，最后一个简答题竟然是论文汇报的内容（每组一篇论文并在课上汇报），这题挺佩服老师的。复习的时候最慌的是“特征选择与提取”这一部分，结果没考，我笑了。一共印了129页的资料，没这资料我可能就…。计算器落在了寝室，在决策树计算时就有些小难受，虽然卷中给出了所有要用到的对数值，但一堆小数乘法以及验算还是花了很多时间，决策树这个计算题是经典的例子。最后，一如既往，考前一罐红牛。

小组中一位大神的CSDN博客：https://blog.csdn.net/weixin\_43756517。

一、名词解释（每个5分，共20分）

监督学习

KNN

线性可分

多层感知器

二、简答题（每个10分，共60分）

两个一模一样的碗，一号碗有30颗水果糖和10颗巧克力糖，二号碗有水果糖和巧克力糖各20颗。现在随机选择一个碗，从中摸出一颗糖，发现是水果糖。请问这颗水果糖来自一号碗的概率有多大？

简述一下k均值聚类的基本思想，并说明如何初始化k

简述一下SVM的基本思想，针对线性不可分问题，SVM有哪些方法

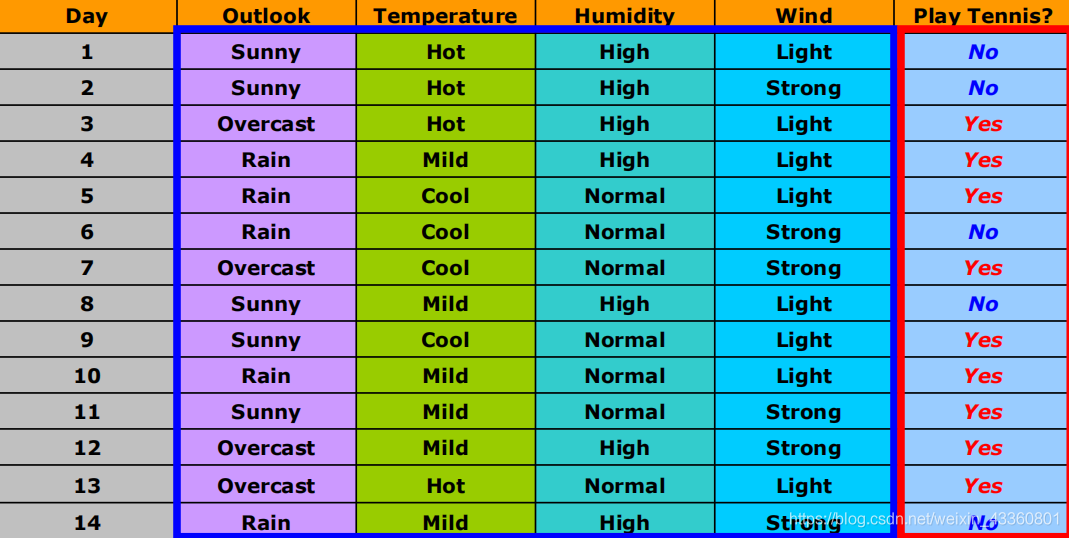
简述一下什么是bagging算法，并比较bagging与AdaBoost的异同

简述一下线性回归，并比较岭回归和lasso回归的区别

从Graph embedding、word embedding、Graph CNN 中选择一个你熟悉的并进行介绍

三、综合分析题（20分）

（个人说明：卷中是10个样本，这里只是一个例子）



（1）从信息增益的角度考虑，Outlook和Humidity哪个作为划分属性，给出具体计算过程

（个人说明：这里给出了可能用到的对数值）

（2）决策树存在过拟合问题，采用什么方法可以减少过拟合风险

————————————————

版权声明：本文为CSDN博主「ALTLI」的原创文章，遵循CC 4.0 BY-SA版权协议，转载请附上原文出处链接及本声明。

原文链接：https://blog.csdn.net/weixin\_43360801/article/details/112060247

山东大学2021-2022机器学习期末考试回忆

本篇博客为大三上机器学习限选课期末考试回忆版，考试过去两天才想起来写，有地方记不太清了。

开卷考试，期末占40分。

2022.1.7

一、名词解释（4\*5’）

1、间隔（margin）

2、KNN

3、激活函数

4、独立同分布

二、简答题（4\*10’）

1、什么是outlier？为什么k-means对outlier敏感？

2、解释一下线性可分。SVM如何处理线性不可分问题？

3、基于最小错误率的贝叶斯分类器 的计算。

4、利用所给数据计算决策树应该选择哪个属性，使用信息增益作为判定标准。

三、（20’）

1、解释训练误差、泛化误差，并画图标出欠拟合、过拟合等。

2、验证集有什么作用？怎样使模型学习到全部数据？

3、为什么梯度下降选择负梯度优化目标函数？

四、（20’）

1、学习率过大或过小有什么后果？如何调整学习率？

2、Adaboost算法中ht的公式推导。

————————————————

版权声明：本文为CSDN博主「番茄炒蛋不加蛋！」的原创文章，遵循CC 4.0 BY-SA版权协议，转载请附上原文出处链接及本声明。

原文链接：<https://blog.csdn.net/m0_52100140/article/details/122398200>

山东大学软件学院2022-2023机器学习期末考试（回忆版）

山东大学软件学院2022-2023机器学习期末考试（回忆版）

前言

1、考试时间：2023/2/18 14:30-16:30（因疫情推迟到开学考的期末考试）

2、考试形式：闭卷（前几届都是开卷，这次闭卷，都不知道怎么复习了/(ㄒoㄒ)/~~ ）

3、考试科目：机器学习基础（老师：XuXinShun）

4、考后感悟：题量比较少，每道题的分值都很高，考试内容大多是算法或模型的原理以及比较优缺点。

5、成绩组成：60%（实验+报告）+40%（期末考试成绩）。（老师说实验查重会零分，不知道会不会这样~ /(ㄒoㄒ)/~）

一、名词解释(25’)

1、测试集

2、Boostrap Samples

3、最大似然估计

4、机器学习

5、KNN

二、简答题(50’)

1、现有三家工厂生产同一个产品，一个产品来自这三家的概率分别为50%，30%，20%，这三家工厂的残次品率分别为0.8%，0.2%，1.2%。那么一个残次品来自第一家工厂的概率为多少。（10’）

2、写出多层感知器的原理，以及BP（反向传播）的基本原理和优缺点。（10’）

3、K-means的基本过程，以及优缺点。（优缺点至少各两个） （10’）

4、介绍Boosting的基本思想，比较Boosting和Bagging的异同。（10’）

5、写出信息增益的公式，写出ID3的优缺点。（10’）

三、综合题(25’)

1、SVM如何解决线性不可分问题。 (5’)

2、写出SVM基本型的形式，写出基本思想及推导过程。（使用ω \omegaω和b） (10’)

3、写出SVM基本型的对偶形式，写出推导过程。（10’）

总结

重点复习各种算法、模型的原理、优缺点和异同，另外，SVM比较重要（查看历年题，发现每次分值都很多，看了一些，然而还是不会/(ㄒoㄒ)/~~）。此次考试考察公式的有信息增益的公式以及SVM基础型和对偶形式的推导。

祝考试顺利~

点个赞呗 ~~

————————————————

版权声明：本文为CSDN博主「南河的南」的原创文章，遵循CC 4.0 BY-SA版权协议，转载请附上原文出处链接及本声明。

原文链接：https://blog.csdn.net/wuxiaoyue2020/article/details/129101326