**概念**

·机器学习是从有限的观测数据中学习出具有一般性的规律，并利用这些规律对未知数据进行预测的方法。

·模式识别专注于研究模式的分类和聚类任务，使用标注样本训练分类器的过程是机器学习的过程。

·人工智能领域：研究如何应⽤计算机来模拟⼈类某些智能⾏为的基本理论、⽅法和技术。

·⼈⼯智能是⼀⻔关于知识的学科，研究如何表示、获取、并使⽤知识的科学。

·模式识别是研究如何使机器具备类似于⼈类对⽬标、⾏为等模式进⾏分析、决策和判别能⼒的理论和⽅法。

·机器学习是指从标注样本中学习其中隐含的规律，并⽤于对未知数据的预测或者分类。

·深度学习是⼀种⼈⼯神经⽹络⽅法，属于机器学习范畴。深度学习⾃适应地从标注样本中学习可辨识的特征。

·计算机视觉是指以计算的⼿段模拟⼈类视觉对视觉信息进⾏表示、分析和理解，属于图像理解的范畴。

·⾃然语⾔处理研究⼈与计算机之间⽤⾃然语⾔进⾏通信的各种理论和⽅法。

名词相关

·模式：具有相同特征的一类事物，有人认为识别的是特征模式或是模式所属类别或同一类中模式的总体。

·样本：研究对象的一个个体，统计学中的实例，观测

·样本集：样本的集合。统计学中的样本指的是样本集

·类/类别：样本所属的子集，同一类样本的特征具有相似性

·特征：对样本的观测，通常用可以度量的值表示

·特征向量：样本的特征向量构成样本的特征空间，空间的维数是特征的个数，一个样本是特征空间中的一个点

·标注样本：事先知道类别标签的样本

·模式识别：利用计算方法根据样本的特征对样本进行类别划分

1. **监督学习与非监督学习：**

**·监督学习（分类）**：已知类别划分，训练样本含类别标签，学习分类器对未标注样本进行决策

识别系统：数据获取->预处理->特征提取与选择->分类器设计与训练->分类决策

判别式：①判别分析：**线性判别函数(Fisher)** & 二次判别函数 ②距离度量分类器：最小距离分类器 & **K近邻分类器**（最近邻分类器k=1） ③最小误差线性判别函数：感知器准则、最小均方准则、最小二乘准则 ④神经网络：感知器网络、线性神经网络、多层前馈型网络（BP）、卷积神经网络 ⑤支持向量机：**线性分类器** & 非线性分类器

PS：K近邻的错误率：n和k->∞，误差率等于贝叶斯误差率（真实概率）

生成式：①**贝叶斯决策**：最小风险贝叶斯决策 & 最小错误率贝叶斯决策 ②概率密度估计：参数估计（极大似然&贝叶斯） & 非参数估计（核密度估计），考查类条件概率为正态（高斯）分布下的贝叶斯决策

**·非监督学习（聚类）**：事先不知道划分的类别，没有标注的样本，通过未标注样本间相似性进行划分

识别系统：数据获取->预处理->特征提取与选择->聚类->结果解释

聚类：①基于样本间距离的聚类：**动态聚类（K均值）** & 层次聚类 ②基于样本概率分布模型的聚类：混合密度聚类（GMM） & 密度聚类（DBSCAN），括号内为经典方法

1. **分类与聚类**

分类是监督学习的一种，聚类是非监督学习的一种。

1. **极大似然估计和贝叶斯估计**

这两种方法是对类条件概率密度函数的估计方法，都属于参数估计方法，估计的准确性由决策分类的错误率评价。极大似然估计将参数θ看做是确定而未知的参数；贝叶斯估计将参数θ看做是随机变量服从某一分布（先验概率分布）p(θ)。

1. **结构风险最小化与经验风险最小化（SRM & ERM）**

ERM：直接学习输入空间到输出空间的映射f，不涉及概率密度估计，建立损失函数，求解最优化问题

SRM：正则化对模型复杂度惩罚，为了避免对有限样本的过拟合。同一个子集中VC维的置信区间相同。在每一个子集中寻找最小经验风险，通常它随着自己复杂度的增加而减小。选择最小经验风险与置信区间之和最小的子集，就可以达到期望风险的最小，这个子集中使经验风险最小的函数就是要求的最优函数

SVM：对权向量的范数约束

1. **特征降维**

特征提取：对原始特征进行变换，生成新的特征，消除或减少特征之间的相关性

特征选择：从已有的特征中挑选对分类最有利的特征子集

目的：特征降维，缓解维度爆炸；特征具有可分性、独立性、数量少的特点

**简答题知识点——基本原理**

Fisher判别：

·将样本投影到一条直线上，使同类样本尽可能狙击，异类样本尽可能分离、在投影上确定分类的阈值。过阈值点且与投影方向垂直的超平面是两类的分类面

·基本思想：准则函数使得类内方差最小，类外方差最大，在这个基础上推导出法向量的公式

·线性判别函数的决策面是超平面

·线性判别函数的意义：f(x) = WtX + w0，w0叫阈值权，决定了分类线的位置，Wt和W0共同决定了线性判别函数。对于函数来说，法向量斜率>0的一边是正侧。

SVM：

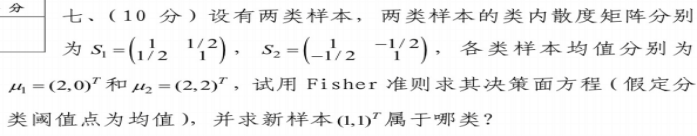
·统计学习理论：期望风险——数学期望、整体均值；经验风险——样本均值。风险也就是代价、损失。

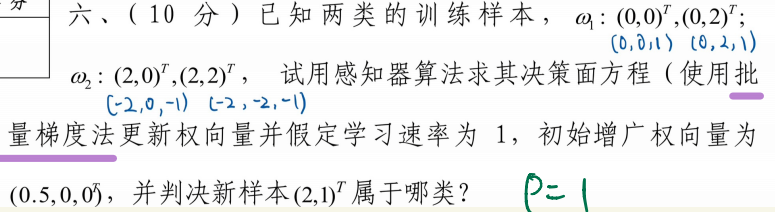
·R(w) = Re(w) + φ(h/n)，Re(w)是拟合误差，后者是VC confidence，代表了模型复杂度，和n成反比，和h成正比，h是衡量函数分类能力的变量，数据集的复杂度越高，h越高 。强调期望风险的上界：经验风险和VC置信（模型的复杂程度和样本数成反比（样本数越小，VC置信区间越大））共同决定。降低VC维度，提高样本数，才能降低VC置信，才能降低期望风险。

·基本思想：分类面在两类训练样本的中间，而且分类间隔最大，提出了最优超平面的思想。

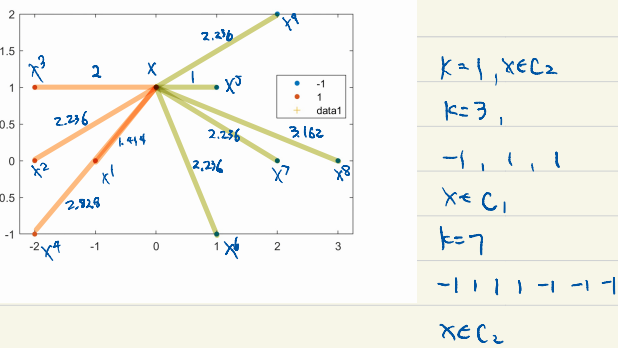
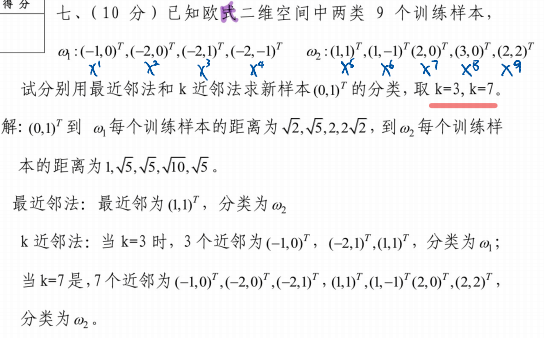
·线性不可分问题：可以通过源空间转换为新空间的限行问题，利用核矩阵和对称半正定矩阵。源空间计算内积，而不是在高位特征空间计算内积

**大题**

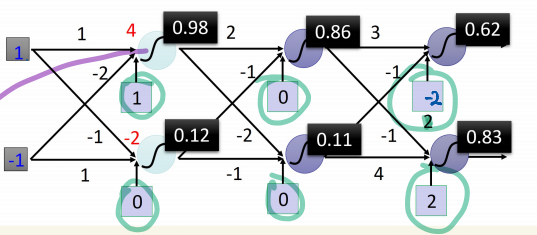
1. Fisher（最大类间方差，最小类内方差）
2. Perception（最小错分准则）



1. K-nearest（分类，距离度量）

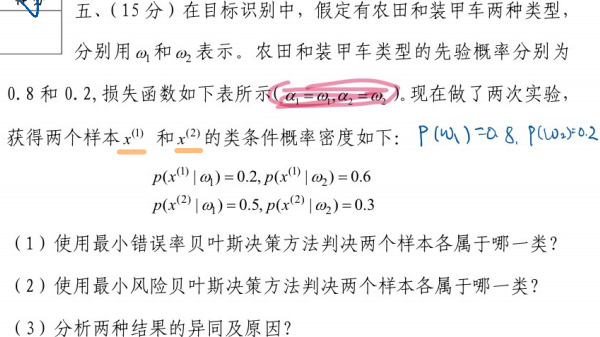


1. 多层前馈网络（必考）

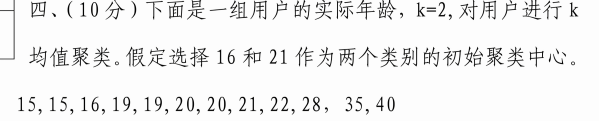


（手写添加）公式和计算说明：强调：神经网络中的偏置使用加法

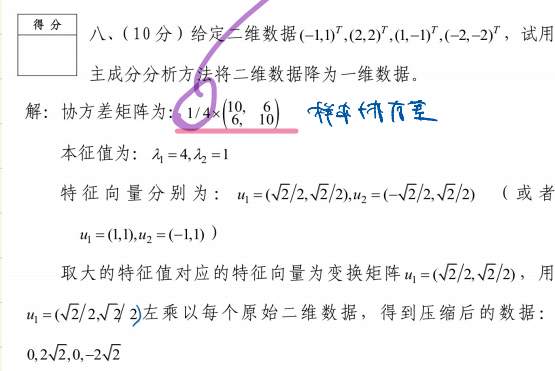
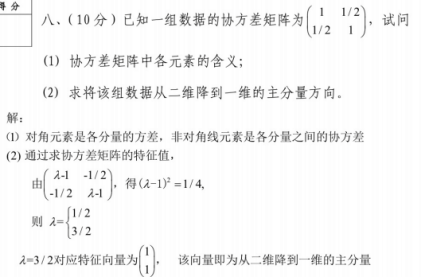
1. Bayes



1. K-means（聚类，距离度量）



1. PCA（降维）



无偏估计左边n换成n-1

特征向量求解方法