模式识别复习

# 模式识别系统构成

## 监督模式识别

* 信息获取与预处理
* 特征提取与选择
* 分类器设计（训练）
* 分类决策

## 非监督模式识别

* 信息获取与预处理
* 特征提取与选择
* 聚类（自学习）
* 结果解释

## 实例

OCR（光学字符识别）是通过扫描仪把印刷或手写的文字稿件输入到计算机中，由计算机自动识别出其中的文字内容。

### 信息获取与预处理

对印刷或手写的文字稿件扫描输入，并将内容图像进行二值化等处理，分割单字。

### 特征提取与选择

将每个单字向各个方向投影，得到像素密度分布；提取笔画分解信息等。

### 分类器设计

OCR问题即多类分类问题，利用大量样本数据，训练多类分类器。并结合已有的对文字结构的认知，提高其准确性。

### 分类决策

根据多类分类器的输出可能结果，结合上下文的联系，得出最终的估计字符识别结果。

# 贝叶斯决策

## 最小风险贝叶斯决策

### 定义

样本

状态空间

决策空间

对实际状态为的向量，采取决策所带来的损失为

对某个样本，属于各个状态的后验概率是，则对他采取决策的期望损失为：

设有一决策规则，他对所有样本决策造成的期望损失为：

最小风险贝叶斯决策即：

由于和非负，且已知，要使积分最小，就要使对所有使最小。

### 步骤

1. 利用贝叶斯公式计算后验概率：
2. 利用决策表，计算条件风险：
3. 选择风险最小的决策：

### 两类问题

### 例题

1. 对两类问题，若损失函数，试求基于最小风险贝叶斯决策分界面处的两类错误率与的关系。

**分界面处**

**因此**

## 最小错误率贝叶斯决策

即最小风险贝叶斯决策的特殊情况：

### 两类问题

## 正态分布时的决策

### 一元正态分布

### 多元正态分布

### 最小错误率贝叶斯决策

判别函数：

决策面方程为：

### 例题

1. 设一个二维空间中的两类样本服从正态分布，其参数分别为，，，，先验概率,试求基于最小错误率的贝叶斯决策分界面方程。

**因为两类样本服从正态分布，其比为**

**分界面处**

所以

**为圆的方程**

# 概率密度函数的估计

## 最大似然估计

### 定义

每类的样本集：

其中的样本都是从密度为中总体中独立抽取出来的。（独立同分布条件）

因此，获得样本集的概率即出现其中各个样本的联合概率：

其为参数相对于样本集的似然函数。

最大似然估计量即：

为了便于分析，定义对数似然估计函数：

### 计算

求解

### 正态分布

令

解得

### 例题

1. 设为来自点二项分布的样本集，即  
   试求参数的最大似然估计量

对数似然估计函数为

对求导有

求极值有

得最大似然估计为

## 贝叶斯估计

### 定义

损失函数：把估计为所造成的损失

期望风险：，其中

条件风险：，其中

贝叶斯估计，即对所有的，最小化条件风险。

常用平方误差损失函数，此时贝叶斯估计量在给定时的条件期望：

在给定样本集下，的贝叶斯估计是：

### 计算

1. 确定先验分布
2. 求样本集的联合分布
3. 求的后验概率分布
4. 求的贝叶斯估计量

也可直接推断总体分布

### 正态分布

假设均值是待估计参数，方差为已知参数，其分布密度为：

假设均值的先验分布也是正态分布，其均值为、方差为，即

利用下式估计

分母用于归一化，暂略

将所有与无关的量写入常数中

其中

所以

### 例题

1. 假定损失函数为二次函数，以及P的先验密度为均匀分布，即，在此条件下，求题3的贝叶斯估计量。

已知Beta分布的期望为，所以

1. 设总体分布密度为，并设，用贝叶斯估计计算。已知的先验分布。

# 线性分类器

## Fisher线性判别分析（LDA）

### 定义

目标是找到一个投影方向，投影后，样本变为

在原样本空间中，类均值向量为

各类的类内离散度矩阵为

总类内离散度矩阵为

类间离散度矩阵为

在投影以后的一维空间，两类的均值分别为

各类的类内离散度（是一个值）为

总类内离散度为

类间离散度为

我们希望最终结果使两类尽可能分开，而各类内尽可能聚集，因此可有如下准则

代入原样本空间的式子，可得

解得

### 例题

1. 设两类样本的类内离散矩阵分别为：  
   ，  
   试用fisher准则求其决策面方程。

## 感知器

### 定义

可以直接得到完整的线性判别函数。

首先将样本向量和权向量增广

线性判别函数为

定义一个新变量（规范化增广样本向量）

此时，样本可分性的条件即存在（解向量）

感知器准则函数为对所有错分样本的求和惩罚

解向量则为

### 计算

梯度下降法

其中

因此，迭代修正公式即

### 例题

1. 用感知器算法求下列模式分类的解向量：  
     
   假设步长。

此时，可以保证所有样本都被正确分类。

所以，

# 非线性分类器

## 人工神经网络

### 定义

神经元接受信号，当信号的加权和大于阈值，则神经元激活，输出信号。

激活函数可以是阶跃函数，但其数学性质不够好，难以建立模型，故使用Sigmoid函数（S形函数）：

其中

即上层神经元的输入加权和。

### 反向传播算法（BP算法）

#### 确定结构

总共层（层为输入层，层为输出层，层为隐层）。

第层有个节点。

#### 选取样本

#### 初始化权值

可以随机取较小的数。

#### 前馈阶段

计算估计计算值

计算估计激活值

前馈计算每一层的估计值。

#### 计算输出层梯度

代价函数

偏导

其中

所以

#### 反向传播

其中

其中的第一项，由下式通过反向传播的层的计算

#### 修正参数

修正后，重复前馈、反向传播和修正参数步骤，直至误差小于设定值，或迭代次数过多。

### 例题

1. 由M-P模型组成的神经元网络的结构与参数如图所示，已知，试问该网络与什么逻辑运算等价。M-C使用的模型参数为：。  
   

显然，只有时，神经元激活。故该网络等价于与运算：AND。

# 特征选择与提取

## 基于类别可分性判据的特征提取

### 判据

判据

### 提取

求矩阵的特征值：，从大到小排序。

选取其最大的个特征值对应的特征向量作为最优变换矩阵。

### 例题

1. 已知有两类数据,分别为：  
      
   试求该数据的类内及类间离散矩阵和，并求使达到最大的特征提取。

特征值求得

因此，选取对应的特征向量

## K-L变换

### 简述

通过选择产生矩阵的前d大的特征值对应的特征向量作为变换矩阵。

当去掉均值信息时，可采用协方差矩阵作为产生矩阵，此时K-L变换等价于主成分分析（PCA）。

### 从类均值提取判别信息

计算总类内离散度

将其作为产生矩阵进行K-L变换，求解特征值和特征向量。

性能指标

利用性能指标计算，从大到小排序，选取前d个特征的特征向量作为变换矩阵。

### 例题

1. 设有一个两类问题，先验概率相等，特征为二维向量，类均值向量分别为  
   协方差矩阵分别是  
   利用K-L变换计算变换矩阵。

所以选

作为变换矩阵。