**1、利用最大似然估计估计高斯分布的均值与方差，并分析优缺点。**

高斯分布的最大似然估计 1.pdf 第24页。

最大似然估计：把待估的参数看作是确定的量，只是其取值是未知的。最佳估计就是使得产生已观测到的样本的概率最大的那个值。

推导公式：

概率公式：

最大似然估计是下面方程组的解：

优缺点：最大似然估计法对任何总体都可以用，从它得到的估计量具有一致性和有效性，即使不具有无偏性，也常常能够修改成无偏估计量。可以证明，在一定条件下，未知参数的最大似然估计与其真值之差可以任意小，所以，从某种意义上说没有比最大似然估计更好的估计。但是，并不是所有的待估计的参数都能求到似然估计量，因为求最大似然估计量时，往往要解一个似然方程（或方程组），有时比较难解或根本就写不出有限形式的解。

**2、阐述两种线性分类器分类的基本原理，如LDA、最小二乘等，并讨论他们之间的区别与联系。**

分类策略 3.pdf 第5页。

（1）Fisher线性判别分析LDA（Linearity Distinction Analysis）：

基本思想：对于两个类别线性分类的问题，选择合适的阈值，使得Fisher准则函数达到极值的向量作为最佳投影方向，与投影方向垂直的超平面就是两类的分类面，使得样本在该方向上投影后，达到最大的类间离散度和最小的类内离散度。

Fisher线性判别并不对样本的分布进行任何假设，但在很多情况下，当样本维数比较高且样本数也比较多时，投影到一维空间后样本接近正态分布，这时可以在一维空间中用样本拟合正态分布，用得到的参数来确定分类阈值。

推导公式：

类间离散度：

类内离散度：

所以有目标函数：

其中

（2）感知机

基本思想：对于线性判别函数，当模式的维数已知时，判别函数的形式实际上就已经确定下来，线性判别的过程即是确定权向量。感知机是一种神经网络模型，其特点是随意确定判别函数初始值，在对样本分类训练过程中，针对分类错误的样本不断进行修正，逐步迭代直至最终分类符合预定标准，从而确定权向量值。可以证明感知机是一种收敛算法，只要模式类别是线性可分的，就可以在有限的迭代步数里求出权向量的解。

推导公式：

优缺点：

优点：简单、便于实现。

缺点：结果不唯一，在线性不可分情况下不收敛。

（3）两者的区别与联系：

Fisher线性判别是把线性分类器的设计分为两步，一是确定最优方向，二是在这个方向上确定分类阈值；感知机则是通过不断迭代直接得到完整的线性判别函数。Fisher线性判别根据阈值选择投影方向达到预期分类效果，而感知机算法因为不是收敛算法，可能不能得到很好的分类结果。

**3、阐述支持向量机的基本原理。**

支持向量机 5.pdf 第6页。

主要思想：针对线性可分情况进行分析，对于线性不可分的情况，通过使用非线性映射算法，将低维输入空间线性不可分的样本转化为高维特征空间样本，使其线性可分，从而使得在高维特征空间采用线性算法对样本的非线性特征进行线性分析成为可能。它基于结构风险最小化理论，在特征空间中建构最优分割超平面，使得学习器得到全局最优化，并且在整个样本空间的期望风险以某个概率满足一定上界。

推导过程：

求解最大的margin值情况下的分类超平面：

数学求解后，根据Karush‐Kuhn‐Tucker (KKT) 条件，分类面为最优分类面的条件为：

决策方程：

**4、以K均值聚类为例，阐述EM算法基本原理。**

K均值算法 8.pdf 第5页。

K均值聚类：给定D维空间上的数据集，并不知道这些数据集所对应的类型和标号，通过聚类方法将这些数据集划分为K类。对于K个聚类中的每一类，分别建立一个代表点，将每一个样本划归到离该样本点最近的代表点所代表的聚类。目的是最小化一个准则函数。

EM算法主要思想：假设我们估计A和B两个参数，在开始状态下两者都是未知的，并且知道了A的信息就可以得到B的信息，反过来知道了B的信息也就得到了A。可以考虑首先赋予A某种初值，以此得到B的估计值，然后从B的当前值出发，重新估计A的取值，这个过程一直持续到收敛为止。

K均值聚类推导：

（1）对于样本，定义一个聚类标注，即如果属于第K个聚类，则

（2）准则函数

（3）两步走策略：

①初始化，按照最优化准则产生

②根据产生的，按照最优准则产生

③根据产生的，按照最优准则产生。

重复②、③过程至收敛。

过程示例：







**5、通过含有一层隐藏层的神经网络推导BP反传算法。**

人工神经网络 4.pdf 第3页。

BP反传算法 4.pdf 第10页。

BP反传算法：三层前馈网络的学习算法比较复杂，其主要困难是中间的隐层不直接与外界连接，无法直接计算其误差。

BP反传算法的基本思想是在训练开始之前，随机地赋予各权值一定的初值。训练过程中，轮流对网络施加各种训练样本。当某个训练样本作用于神经网络输入端后，利用当前权值计算神经网络的输出，这是一个信号从输入到隐藏层再到输出层的过程，称为前向过程。

考察所得到的输出与训练样本的已知正确输出之间的误差，根据误差对这些输出层节点权值的偏导修正这些权值；再将误差反向传递到隐藏层各节点上，根据误差对这些隐藏层节点权值的偏导修正这些权值，这个过程称为反向过程。

不断地重复进行前向过程与反向过程，直到在一次训练中误差水平达到预先设定的阈值，或者训练时间达到了设定的上限。

两层前馈神经网络：



三层神经网络：



对于输出层而言：

对于隐藏层而言：

BP反传算法总结：

（1）初始化权重；

（2）对于输入的训练样本，求取每个节点输出和最终输出层的输出值；

（3）对输出层求取；

（4）对于隐藏层求取；

（5）求取输出误差对于每个权重的梯度；

（6）更新权重。

（7）重复进行（2）~（6）步进行迭代至收敛。

评价：

用梯度法求非线性函数极值，因而有可能陷入局部极小点，不能保证收敛到全局极小点；算法收敛速度很慢；新加入的样本会影响已有样本产生的学习结果。

权值初始值的选择：如果权值初始值都为零或都相同，隐层各单元不能出现差异，运算不能正常进行。因此，通常用较小的随机数（例如在-0.3～0.3之间的随机数）作为权值初始值。初始值对收敛有影响，当计算不收敛时，可以改变初始值试算。

**6、简述Discrete Adaboost的基本原理，并探讨它在某一领域的应用（已经学过的除外）。**

Discrete Adaboost 13.pdf 第9页

核心思想：它是一种迭代算法，其核心思想是针对同一个训练集训练不同的分类器（弱分类器），然后把这些若分类器集合起来，构成一个更强的最终分类器（强分类器）。其算法本身是通过改变数据分布来实现的，它根据每次训练集中每个样本的分类是否正确，以及上次总体分布的准确率，来确定每个样本的权值，将修改过权值的新数据集送给下层分类器进行训练，最后将每次训练得到的分类器最后融合起来，作为最终的分类器。使用Adaboost可以排除一些不必要的训练数据特征，并主要集中于关键数据。

训练过程：算法中不同的训练集是通过调整每个样本对应的权重来实现的。开始时，每个样本对应的权重是相同的，在此样本分布下训练出一弱分类器。对于分类错误的样本，加大其对应的权值；而对于分类正确的样本，降低其对应的权值，这样分错的样本就被突出出来，从而得到一个新的样本分布。在新的样本分布下，再次对弱分类器进行训练，得到另一个弱分类器。以此类推，经过T次循环，得到T个弱分类器，把这T个弱分类器按照一定的权值叠加（boost）起来，得到最终需要的强分类器。

三种方式：

（1）有T个弱分类器，产生强分类器H strong。

（2）有T个弱分类器，其各自的权重是，产生强分类器。

（3）进一步考虑弱分类器和样本进行自适应：Adaptive Boosting，即Adaboost。

Adaboost算法通过从大量的特征中挑出最优的特征，并将其转化为对应的弱分类器进行分类使用，从而达到对目标进行分类的目的。

**7、对一幅带噪声的二值图像，利用马尔科夫随机场进行降噪（可能会换成相似的问题）。**

马尔科夫随机场 7.pdf 第5页。

图像去噪 7.pdf 第10页。

马尔科夫随机场：马尔科夫一般是马尔科夫性质的简称，它指的是一个随机变量序列按时间先后关系依次排开的时候，第N+1时刻的分布特性，只与N时刻有关，且与N时刻以前的随机变量的取值无关。随机场包含两个要素：位置和相空间，当给每一个位置按照某种分布随机赋予相空间的一个值之后，其全体就叫做随机场。

马尔科夫随机场在二值图像降噪中的运用主要是指在对图像进行分析时，认为图像上的每个点是否是噪点只与它周围的若干个点有关，而与整体图像无关。所以针对每个点，若它对比周围各点显现出颜色突变，则判定为噪点。算法用能量函数来描述这种对比关系，若将该点颜色改变为另一个值后能量降低，则判定该点为噪点并维持改变后的颜色；反之则判定为非噪点，并将改回原有颜色。

**

能量函数：

梯度下降法：

（1）初始化：；

（2）一次尝试改变一个：；

（3）如果能量降低则改变，如果能量升高则不改变；

（4）依次进行。

评价：

很高效，但只能找到局部极值。

**8、其他问题以Surprise的形式出现。**