

名词解释：

1 样本：对任一个具体的事物，在这门课中都称为一个样本，它是一类事物的一个具体体现，它与模式这个概念联用，则模式表示一类事物的统称，而样本则是该类事物的一个具体体现。

2 模式：英语是 pattern，表示一类事物，如印刷体 A 与手写体 A 属同一模式。B 与 A 则属于不同模式，而每一个具体的字母 A、B 则是它的模式的具体体现，称之为样本。因此模式与样本共同使用时，样本是具体的事物，而模式是对同一类事物概念性的概况。一个人的许多照片是这个人的许多样本，而这个人本身是一个模式。

3 模式类：这个词与模式联合使用，此时模式表示具体的事物，而模式类则是对这一类事物的概念性描述。

4 模式识别：人们在见到一个具体的物品时会分辨出它的类名，如方桌与圆桌都会归结为是桌子。这是人们所具有的认识事物的功能，在这门课中就称为是模式识别。具体的说是从具体事物辨别出它的概念。这门课讨论的是让机器实现事物的分类，因此由机器实现模式识别。这门课就是讨论机器认识事物的基本概念、基本方法。

5 分类器：用来识别具体事物的类别的系统称为分类器

6 模式识别系统：用来实现对所见事物（样本）确定其类别的系统，也称为分类器。

7 特征：一个事件（样本）有若干属性称为特征，对属性要进行度量，一般有两种方法，一种是定量的，如长度、体积、重量等，可用具体的数量表示，但也可用粗略的方法表示，如一个物体可用“重”、“轻”、“中等”表示，前种方法为定量表示，而后种方法则是定性表示。重与轻变成了一种离散的，或称符号性的表示，它们在数值上有内在的联系。在本门课中一般偏重定量的表示。

8 特征向量：对一个具体事物（样本）往往可用其多个属性来描述，因此，描述该事物用了多个特征，将这些特征有序地排列起来，如一个桌子用长、宽、高三种属性的度量值有序地排列起来，就成为一个向量。这种向量就称为特征向量。每个属性称为它的一个分量，或一个元素。

9 维数：一个向量具有的分量数目，如向量 $X = (x_1, x_2, x_3)^t$ ，则该向量的维数是 3。

$$X = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

10 列向量：将一个向量的分量排列成一列表示，如

11 行向量：将一个向量的分量排列成一行表示，如 $X = (x_1, x_2, x_3)$

12 转置：将一个列向量写成行向量的形式的方法就是转置。如定义 X 为列向量，则 X^T 就是该向量的行向量表示。转置的概念与矩阵中转置的概念一样。

13 特征空间：一种事物的每个属性值都是在一定范围内变化，修改桌子高度一般在 0.5 米到 1.5 米范围内变化，宽度在

0.6 到 1.5 米范围内变化。长度是 1 米到 3 米范围内变化，则由这三个范围限定的一个三维空间就是桌子的特征空间。归纳起来说所讨论问题的特征向量可能取值范围的全体就是特征空间。

14 分类决策：根据一个事物（样本）的属性确定其类别，称为分类决策。

15 分类决策方法：对一事物进行分类决策所用的具体方法，例如一个人身高超过 1.8 米，就判断它是个男人，身高超过 1.8 米就是具体的分类决策方法。

16 学习：让一个机器有分类决策能力，就需要找到具体的分类决策方法，确定分类决策方法的过程统称为学习，就像人认识事物的本领的获取与提高都是通过学习得到的。在本门课中将学习分成有监督学习与无监督学习两种不同的方法。

17 训练：一般将有监督学习的学习方法称之为训练。

18 训练（样本）集：在训练过程中使用的样本集，该样本集中的每个样本的类别已知。例如训练一个将男女进行分类的系统的训练集，应包含一个男生集及一个女生集，这两个集中每个成员的性别是已知的。

19 有监督学习方法：从不同类的训练集数据中体现出的规律性进行分析，从而确定分类决策方法，这种学习方法是在训练集指导下进行的，就像有教师来指导学习一样，称为有监督学习方法。与之相对的是无监督学习方法。

20 无监督学习方法：在一组数据集中寻找其规律性的过程称为无监督学习方法。例如分析数据集中的自然划分（聚类）；分析数据集体现的规律性，并用某种数学形式表示（数据拟合）；分析数据集中各种分量（描述量，特征）之间的相关性（数据挖掘，知识获取）等，这种学习没有训练样本集作指导，这是与有监督学习方法的不同点。

21 先验概率：根据大量统计确定某类事物出现的比例，如我国理工科大学男女生比例大约为 8:2，则在这类学校一个学生是男生的先验概率为 0.8，而为女生的概率是 0.2，这两类概率是互相制约的，因为这两个概率之和应满足总和为 1 的约束。

22 类分布概率密度函数：同一类事物的各个属性都有一定的变化范围，在这些变化范围内的分布密度用一种函数形式表示，则称为类分布概率密度函数（类条件概率密度），这种函数可以是一些著名的普遍运用的函数形式，如正态分布，也可能是更复杂的无法用分析式表示的函数。这种分布密度只对同一类事物而言，因此与其它类事物没有关系。例如，男女生比例是男生与女生这两类事物之间的关系，而男生高度的分布则与女生的分布无关。为了强调是同一类事物内部，因此这种分布密度函数往往表示成条件概率的形式。例如 X 表示某一个学生的特征向量，则，男生的分布概率密度表示成 $P(X|\text{男生})$ ，女生的表示成 $P(X|\text{女生})$ 这两者之间没有任何关系。即一般的情况下 $P(X|w_1)+P(X|w_2) \leq 1$ 可为从 0~1 之间的任意值。

23 后验概率：一个具体事物属于某种类别的概率，例如一个学生用特征向量 X 表示，它是男性或女性的概率表示成 $P(\text{男生}|X)$ 和 $P(\text{女生}|X)$ 这就是后验概率。由于一个学生只可能为两个性别之一，因此有 $P(\text{男生}|X)+P(\text{女生}|X)=1$ 的约束，这一点是与类分布密度函数不同的。后验概率与先验概率也不同，后验概率涉及一个具体事物，而先验概率是泛指一类事物，

因此 $P(\text{男生}|X)$ 和 $P(\text{男生})$ 是两个不同的概念。

24 贝叶斯公式：两个事物 X 与 w 联合出现的概率称为联合概率，可写成 $P(X, w)$ 而它们又可与条件概率联系起来，即 $P(X, w)=P(X|w)P(w)=P(w|x)P(w)$ 这就是贝叶斯公式。如果将上式中各个项与先验概率，类分布密度函数以及后验概率联合起来，可以找到利用先验概率，分布密度函数计算后验概率的方法。

25 贝叶斯决策理论：根据先验概率、类分布密度函数以及后验概率这些量来实现分类决策的方法，称为贝叶斯决策理论。由于这些量之间符合贝叶斯公式，因此称为贝叶斯决策理论。

26 基于最小错误率的贝叶斯决策：根据一个事物后验概率最大作为分类依据的决策，称为基于最小错误率的贝叶斯决策。从统计上讲，即从平均错误率角度看，分类错误率为最小，因此称为基于最小错误率的贝叶斯决策。

27 风险决策：对事物进行分类或做某种决策，都有可能产生错误，不同性质的错误就会带来各种不同程度的损失，因而作决策是要冒风险的。考虑到决策后果（风险）的决策是风险决策。如进行股票交易要冒风险，投资，确定建设项目，规划等都要冒风险，在衡量了可能遇到的风险后所作的决策称为风险决策。

28 基于最小风险的贝叶斯决策：如果样本 X 的实际类别为 w_i ，而作决策为 j 则可以定义此时作 j 决策的风险为 $r(j|w_i)$ ，由此可以确定对样本 X 做 j 决策的期望损失，比较做不同决策的期望损失，选择期望损失最小的决策后最终决策。就是基于最小风险的贝叶斯决策。

29 判别函数：是一组与各类别有关的函数，对每一个样本可以计算出这组函数的所有函数值，然后依据这些函数值的极值(最大或最小)做分类决策。例如基于最小错误率的贝叶斯决策的判别函数就是样本的每类后验概率，基于最小风险的贝叶斯决策中的判别函数是该样本对每个决策的期望风险。

30 决策域与决策面：根据判别函数组中哪一个判别函数值为极值为准则可将特征空间划分成不同的区域，称为决策域，相邻决策域的边界是决策分界面或称决策面。例如两类问题的基于最小错误率的贝叶斯决策将整个特征空间划分成两个决策域，在同一个决策域中的每一点由同一类的后验概率占主导地位。

31 参数估计：使用贝叶斯决策要知道先验概率，类分布密度函数等统计参数，为此，要从训练样本集中估计出这些统计参数，这就是参数估计。

32 非参数估计：在分布密度函数形式也不确定条件下，估计统计参数，称为非参数估计。

33 非参数分类器：不以统计参数为分类决策依据的分类决策方法称为非参数分类器，线性分类器、非线性分类器以及近邻分类器都属于这种分类器，它们不需要统计参数。

34 线性分类器：判别函数为线性函数的分类器是线性分类器，此时决策分界面的方程是线性方程。

35 非线性分类器：是非参数分类器的一种，其中判别函数或决策面方程是某种特定的非线性函数，如二次函数，多项

式函数等。

36 分段线性分类器：相邻决策域的界面用分段线性函数表示的分类器。

37 近邻法：通过计算待分类样本与已知类别的模板样本集计算相似度（相邻性），从而以最相似模板样本的类别作为分类依据的方法。

38 K-近邻法：是近邻法中的一种，对待分类样本找到 K 个近邻，并以该 K 个近邻中的主导类别作为待分类样本的分类依据，当 K=1 时称为最近邻法。

39 Fisher 准则判别函数：线性分类器中的一种分类决策面设计方法，是由 Fisher 提出而得名，一般用于两类别分类器中。该种设计方法要找到分界面的最佳法线，使两类别训练样本到该法线向量的投影体现“类间尽可能分离，类内尽可能密集”的最佳准则。

40 感知准则函数：是线性分类器的另一种著名设计方法。该方法通过迭代优化确定最佳分界面。最佳准则取决于所使用的最佳准则，如最小错分数准则等。其特点是利用错分类信息对当前的分界面进行修正。

41 感知器：使用感知准则函数设计的分类器称为感知器，它是神经网络中最简单的一种，是人工神经前期研究的成果。

42 剪辑近邻法：对近邻法使用的模板样本集通过剪辑进行修正，以达到进一步减小错误率，压缩模板样本数量为目的。

43 压缩近邻法：是另一种改进近邻法的方法，以最大限度削减近邻法中模板样本数量为目的。

44 支持向量机：是基于统计学习理论的一种分类器设计方法，是近年来在理论及实际问题都有重大影响的一种新方法。就分类器设计而言，它以设计线性分类器为基础，扩展到非线性分类器。在设计线性分类器时，又分线性可分以及线性不可分两种情况。在线性可分条件下，即两个类别训练样本集可用线性分界面无错误分开的条件下，找到使两类别训练样本正确分类的一个最佳分界面。最佳条件是指两类样本最靠近分界面的样本（称为支持向量）到该分界面的法向距离最大。也就是说使分界面两侧形成的一个隔离带（带中没有任一类训练样本）的间隔最宽。在线性不可分条件下，即两类样本无法用线性界面无错分开的条件下，最佳准则改为综合考虑对错样本进行控制，与使间隔带尽可能宽这两个条件。

对线性不可分条件下分类也可使用非线性分类器，支持向量机中采用将原特征空间，用非线性映射到一个新空间，并在该空间采用线性分类器的方法。

45 特征选择：对样本采用多维特征向量描述，各个特征向量对分类起的作用不一样，在原特征空间中挑选中部分对分类较有效的特征组成新的降维特征空间，以降低计算复杂度，同时改进或不过分降低分类效果特征选择的另一种含义是指人们通过观察分析选择适用于分类的特征。

46 特征提取：特征提取是从样本的某种描述状态（如一幅具体的图象，一段声波信等）提取出所需要的，用另一种形

式表示的特征（如在图象中抽取轮廓信号，声音信号提取中不同频率的信息等）。这种提取方法往往都要通过某种形式的变换。如对原特征空间进行线性变换或其它变换，滤波也是变换的一种形式。特征提取也往往可以达到降维的目的。目前使用什么样方法提取特征。主要靠设计人员确定，如选择什么样的变换，主要由人来决定，但如确定用某种线性变换，则线性变换的参数可通过计算来确定。

47 K-L 变换：K-L 变换是一种特殊的正交变换，它是通过对样本集协方差矩阵求的特征值与特征向量的方式构造正交变换。利用部分特征值最大的特征向量构造的正交变换可对原信号进行降维重构，重构后的信号与原信号之差为截尾误差时的最佳正交变换。最佳条件是指在降维数相同条件下，K-L 变换的平均截尾误差平方和比任何一个其它正交变换要小。K-L 变换的这种性质对信息压缩有价值，在模式识别中广泛用于特征提取。

48 主分量分析：主分量分析是 K-L 变换的另一种名称。

49 聚类：一个数据集可能由若干个聚集成群的子集，每个子集称为一个聚类。找出这些按自然分布的聚类，是聚类算法的目的与任务。聚类方法一般可分为动态聚类方法与分级聚类方法两大类。

50 动态聚类方法：通过迭代使聚类划分逐步优化的方法，典型的方法有 C-均值算法，ISODATA 算法等，由于动态聚类方法是一种迭代优化算法，需要确定一种准则函数，迭代过程是使准则函数值趋于极度值的过程。准则函数趋于极值应能反映聚类趋于更合理。

51 C-均值算法：是动态聚类方法中的一个典型方法。其目的是将一数据集，按自然密集程度划分成 C 个聚类，它的准则函数是对所有 C 个聚类中每个数据到其各自均值的距离平方和的总和为最小。计算距离的最简单形式是欧式距离。但也可使用其它形式的距离。迭代过程是计算这个数据，从现属聚类转移至其它聚类，是否能使准则函数值减小为依据，将该数据转移至合适聚类，直至这种数据转移不再发生为止。在数据转移过程中各个聚类的均值也随之改变。

52 ISODATA 算法：是另一种典型的动态聚类方法，它与 C-均值算法的主要不同点是它包含聚类的分裂与合并过程，从而可以根据需要改变聚类的数目。

53 分级聚类方法：对数据集采用逐级合并的方法进行聚类，在初始时整个数据集的每个数据自成一类，然后按相似度最高的要求进行合并，随着相似度要求逐次降低，小的集群逐级合并，聚类数量逐渐减少。这种方法基于分类学原理，如人与类人猿相近，猫与虎同属猫科，就是基于这种原理。

54 人工神经元模型：1940 年 McCulloch 与 Pitts 提出仿造神经元的人工神经元模型。神经网络的输入端接收多维输入向量（如模式识别中常用的特征向量），输出端只有一个。该模型由两部分功能组成，一部分功能将输入端信号进行线性迭加，另一部分是对该迭加值进行非线性映射及从输出端输出。

McCulloch-Pitts 模型中的非线性映射是阈值函数，输入值从负无穷到某个值内输出为 0，输入一旦达到该阈值，输出为 1。这种 0—1 输出也可改成 -1—+1 输出。近来人们根据需要，在人工神经元模型中也采用其它类型的非线性映射。

55 McCulloch-Pitts 模型：就是 McCulloch 和 Pitts 俩人在 1940 提出的人工神经元模型。

56 单层感知器：一个 McCulloch-Pitts 模型可实现的功能是两类别线性分类器，输入向量的线性求和是判别函数，非线性映射及输出的 0、1 值表示了分类结果。这种线性分类器的系数若采用感知准则函数方法确定，它就是一个感知器。这种感知器与多层感知器相比，只有一层，因而称为单层感知器。

57 人工神经网络：将多个人工神经元模型连接成网就是人工神经网络。常用的人工神经网络有多层感知器，Hopfield 模型等。

58 多层感知器：将多个感知器组成分层结构，由输入结点层向第一层结点传递信息，而较低层结点输出的信息往更高一层传递。这种信息单向传递的层次结构称为多层感知器。多层感知器可以实现复杂的逻辑运算功能，输入与输出之间的非线性映射功能。

59 前馈网络：前馈网络是指像多层感知器这样的信息单向传输的网络。

60 隐层结点：由两层及两层以上结点组成的前馈网络中处于非输出层的结点称为隐层结点，一个只含一个隐含层的前馈网络称为双层前馈网络。含两个隐含层的称三层前馈网络。一般使用至多两个隐含层的网络。

61 BP 算法：又称误差回传算法，是在感知准则函数训练方法基础上发展的方法。

62 网络的期望输出值：前馈网络在输入某个向量时期望的输出值。

63 网络的实际输出值：前馈网络在输入某个向量时输出端的实际输出值。

64 误差回传算法：前馈网络中各种参数的设定是通过训练样本集迭代的训练过程确定的。当网络中的参数不合适时，输入向量的实际输出与期望输出值之间不一致，称为误差。利用这种误差从输出层向输入层逐级反传，以确定各个参数的修正量，这种方法因而称之为误差回传算法。

65 径向基函数网络：在多层感知器的人工神经元一般采用 S 型函数作非线性映射。而径向基函数网络则采用单峰型函数（如高斯函数）为非线性映射函数。在实现数据拟合等等应用中较多。训练也较简单。

66 Hopfield 模型：是人工神经网络中一个典型网络，该网络中的所有神经元实行全联接，即任一个神经元的输入输出都联接至其它结点的输入端，因此每一个结点的状态对其它结点的状态都有牵扯作用。该网络的联接数值由 Hebb 规则确定。其功能可实现联想记忆，是人工神经网络中的一种动力系统。网络具有存储模式功能，有自己的稳态，其原理与电子线路中的双稳态触发器相仿。

67 Hebb 规则：Hebb 是神经生理学家，在 1949 年提出一种假设，认为脑细胞中突触强度根据其感受到的经验改变。他因此提出这种突触强度的改变与神经在触发前后突触之间的相关成正比。由于在 Hopfield 模型中采用的联接规则与 Hebb 的修改相比，因而在人工神经网络中广泛使用的规则以 Hebb 冠名。这种规则除了在 Hopfield 模型中使用外，在非监督学习方法的人工神经网络中广泛采用。

68 非监督 Hebb 学习方法：人工神经网络也用来分析数据集内的一些规律，如方差分析，主分量分析等，由于其训练方法采

用了 Hebb 规则，因而称为非监督 Hebb 学习方法。

69 非监督竞争学习方法：在非监督 Hebb 学习方法中引入竞争网络，竞争网络在输出结点层内引入侧抑制机制。每个输出结点给自身以自增强信息，而对其它结点输入侧抑制信息。从而达到“胜者拥有一切”的目的。有了竞争机制，人工神经网络可以实现聚类等功能，一些典型的网络模型如自适应共振理论（Adaptive Resonance Theory），自组织特征映射等都属于具有竞争功能的非监督学习人工神经网络。

70 自适应共振理论：自适应共振理论是 Carpenter 和 Grossberg 在 1987 年提出的网络模型，它主要模仿人积累知识的过程。例如人们可以辨别当前见到的事物，是已往见到过的事物，还是从未见到过的事物，如是没有见过的事物，则将其印入脑海中。自适应共振理论模型在网络内部设有竞争网络，计算当前输入信息与已存模型中哪个最相似。如最相似度超过一定阈值，则达到识别模式的目的。如当前输入信息与所有已存模型都不够相似，则判定为新模式，并为其建模与存储。

71 自组织特征映射（Self Organization Feature mapping SOF）是 Kohonen 提出的一种将高维空间中的数据映射到低维空间的一种神经网络，而实现映射的同时要保持数据在原高维空间的相似性。这是指原数据在高维空间较相近，则它们的映象在低维空间中也应相近。这种网络结构的特点是输出结点数量众多，并按一定规则排列。它们接受高维信号，并在某个输出结点中得到高响应。输入高维信号相近，则高响应的输出结点也应相近。