# 提交实验报告的格式要求

实验作业是本课程的重要组成部分，实验报告要简捷、明了，要有统一而简单的格式。下面是实验报告的参考格式要求：

**封面**. 要求排版美观，内容包括：

- 实验名称（标题）
- 实验编号
- 签名
- 姓名
- 学号
- 截止提交日期
- 摘要 (不要超过 1/2 页)

**技术论述**. 约 1~4 页. 包括所使用的技术，如果在实验中涉及到公式的话，则要包括在实验中用到的主要公式。

**实验结果讨论**. 1~2 页. 包括实验中主要生成的客观数据的分析，以及对任何生成的图表、图形等的清楚、明确的介绍和说明。

**实验结果**(页数不限). 包括实验中产生的所有图表、图形。每幅图表、图形要单独编号，以利于在讨论过程中方便引用。

**参考文献**(页数不限). 列出报告中所引用的文献，若为书，则要列出书中的页码。

**附录**. 程序清单. 学生自己编写的所有程序的清单。对于公认的有名称的标准例程，其源码程序不需要打印出清单，但需要注明出处。

**实验报告版面要求**. 整个报告打印在标准 A4(21 x 29.7 cm)纸张上，报告用订书机左侧装订。

**编程中注意的事项：** 目前有很多可以完成本课程实验中的某些函数的现成程序包。但是在本课程的实验中，若整个实验过程仅仅用已经存在的、现成的例程去实现是不允许的，也不会有高分的。

下一页是本课程试验报告的封面样例

## Proj01-01：模式类实验数据的生成

在 Matlab 中提供了很多产生随机数和随机向量的函数，以及计算随机函数的概率密度值的函数。下面是几个较常用的函数：

rand() 生成均匀分布随机数；　randn() 生成高斯分布随机数

mvnrnd() 生成多元高斯分布的随机向量矩阵

mvnpdf() 计算多元高斯分布的概率密度函数值

认真阅读上述函数 Matlab 的在线帮助，以及 Matlab 中的绘图函数。

完成下面的实验：

(1) 在一维区间[10,70]中，生成 1000 个均匀分布的随机数，然后统计并绘制这些数的直方图；在二维区间[1,5]X[20,30]中，生成 5000 个均匀分布的二维随机点，并绘制出它们的二维散点图；在三维区间[10,50]X[30,60]X[10,15]中，生成 10000 个均匀分布的三维随机点量，并绘制出它们的三维散点图。

(2) 生成两组各 1000 个具有不同均值和方差的一维高斯分布的随机数，然后统计并绘制这些点的直方图；生成三组各 1000 个具有不同均值矢量和协方差矩阵的二维随机矢量，并绘制出它们的二维散点图；生成五组各 1000 个具有不同均值矢量和协方差矩阵的三维随机矢量，并绘制出它们的三维散点图。进一步，绘制上述三维随机矢量数据集合的二维投影散点图。可以指定模式向量的其中两个分量，将集合中每个向量的这两个分量提取出来构成一个 2 维模式子分量的向量集合，然后在二维平面上画出该子分量集合的二维散点图。

(3) 确定一个二维的均值矢量和协方差矩阵，然后利用 matlab 中的 meshgrid 函数生成一个二维网格，利用 mvnpdf 函数计算在每个网格点上的概率密度函数值，并绘制出这些函数值的三维曲面图。

## Proj02-01：最小错误率贝叶斯分类器

### （1）基本概念介绍

贝叶斯公式：

$$P(\omega_j|\mathrm{x})=\frac{P(\mathrm{x}|\omega_j)P(\omega_j)}{P(\mathrm{x})}$$

最小错误率贝叶斯分类器：

考虑 c 类样本，令分类器函数为：

$$g_j(\mathrm{x})=P(\mathrm{x}|\omega_j)P(\omega_j) \qquad j=1,2,\cdots,\mathrm{c}$$

则

$$\mathbf{x}\in\omega_i,\quad if \quad i=\arg\max_j\left\{g_j(\mathbf{x})\middle| j=1,2,...,c\right\}$$

称上述为最小错误率贝叶斯分类器。

若样本 x 为 $d$ 维向量，第 $j$ 类 $\omega_j$ 样本的条件概率密度服从均值为 $\mathbf{m}_j$，协方差为 $\mathbf{S}_j$ 的多元高斯分布：

$$P(\mathrm{x}|\omega_j)=\frac{1}{(2\pi)^{d/2}\sqrt{|\mathbf{S}_j|}}e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m}_j)^T\mathbf{S}_j^{-1}(\mathbf{x}-\mathbf{m}_j)}$$

上式中，$|\mathrm{S}_j|$ 是 $\mathrm{S}_j$ 的行列式函数，$T$ 是矩阵转置符号，$\mathrm{S}_j^{-1}$ 是 $\mathrm{S}_j$ 的逆矩阵。上面的高斯概率密度函数值的计算可以使用 matlab 中的函数 mvnpdf() 实现。

### （2）基本实验

编程实现一个可以对两类模式样本进行分类的贝叶斯分类器，假设两个模式类的条件概率分布均为高斯分布。模式类 1 的均值矢量 $\mathrm{m}_1=(1,3)^T$，协方差矩阵为 $\mathrm{S}_1=(1.5,0;0,1)$，模式类 2 的均值矢量 $\mathrm{m}_2=(3,1)^T$，协方差矩阵为 $\mathrm{S}_2=(1,0.5;0.5,2)$，先验概率 $P(\omega_1)=P(\omega_2)=0.5$。

(a) 利用 proj01-01 中的函数为两个模式类各生成 100 个随机样本，并在一幅图中用不同的符号画出这两类样本的二维散点图；

(b) 利用贝叶斯分类器，对(a)中的 200 个样本进行分类，统计正确分类的百分比，并在 2 维图上用不同的颜色画出正确分类和错分的样本；

(c) 若先验概率 $P(\omega_1)$=0.4， $P(\omega_2)$= 0.6，重新进行(b)中的实验。

(d) 对上述实验结果进行分析说明。

## （3）拓展实验

(e) 若在基本实验(2)中，协方差矩阵不变，但类均值向量分别变为 $\mathbf{m}_1 = (1, 3)^T$，$\mathbf{m}_2 = (2, 2)^T$，重新进行上面的实验(a)、(b)、(c)；

(f) 若在基本实验(2)中，协方差矩阵不变，但类均值向量分别变为 $\mathbf{m}_1 = (1, 3)^T$，$\mathbf{m}_2 = (4, 0)^T$，重新进行上面的实验(a)、(b)、(c)；

(g) 若在基本实验(2)中，两个类的均值向量不变，但协方差矩阵分别变为 $S_1$ =(1.5，1；1，1)， $S_2$ =(1，0.5；0.5，2)，重新进行上面的实验(a)、(b)、(c)

(h) 对上述实验结果分别进行分析说明。

## Proj02-02：最小马氏距离分类器

### （1）基本概念介绍

考虑一个由 $d$ 维向量 $\mathbf{x}$ 组成的样本集合 X，若该集合中的样本 $\mathbf{x}$ 服从均值为 $\mathbf{m}$，协方差为 $\mathbf{S}$ 的多元高斯分布：

$$P(\mathbf{x})=\frac{1}{(2\pi)^{d/2}\sqrt{|\mathbf{S}|}}e^{-\frac{1}{2}(\mathbf{x}-\mathbf{m})^T\mathbf{S}^{-1}(\mathbf{x}-\mathbf{m})}$$

则

$$D(\mathbf{x},\mathbf{m})=\sqrt{(\mathbf{x}-\mathbf{m})^T\mathbf{S}^{-1}(\mathbf{x}-\mathbf{m})}$$

称 $D(\mathbf{x},\mathbf{m})$ 为向量 $\mathbf{x}$ 到样本集合 X 的**马氏距离**（Mahalanobis Distance）。

假设服从多元高斯分布的样本集合 X 中有 N 个样本，第 k 个样本为 $\mathbf{x}(\mathrm{k})\in\mathbf{X}$，则样本集合 X 的均值 $\mathbf{m}$、协方差 $\mathbf{S}$ 的估计值由下面的公式给出：

$$\mathbf{m}=\frac{1}{N}\sum_{k=1}^{N}\mathbf{x}(k),\qquad \mathbf{S}=\frac{1}{N}\sum_{k=1}^{N}(\mathbf{x}(k)-\mathbf{m})(\mathbf{x}(k)-\mathbf{m})^T$$

其中，$\mathbf{x}(k)=(x_1(k),x_2(k),...,x_d(k))^T$，$\mathbf{m}=(m_1,m_2,...,m_d)^T$ 均为 $d$ 维向量，S 为 $d\times d$ 的矩阵。

考虑 c 类样本类 $\omega_j$，$j$ =1,2,…,c，均服从均值为 $\mathbf{m}$(j)与协方差矩阵为 $\mathbf{S}$(j)的多元高斯分布。令分类器函数为：

$$g_j(\mathbf{x})=-D(\mathbf{x},\mathbf{m}(j))=-\sqrt{(\mathbf{x}-\mathbf{m}(j))^T\mathbf{S}^{-1}(j)(\mathbf{x}-\mathbf{m}(j))}\;,\qquad j=1,2,\ldots,c$$

则

$$\mathbf{x}\in\omega_i,\quad if\quad i=\arg\max_j\{g_j(\mathbf{x})\,|\,j=1,2,...,c\}$$

称上述为最小马氏距离分类器。

## （2）基本实验

表格 1

| 样本 | $w_1$ | | | $w_2$ | | | $w_3$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ |
| **1** | -5.01 | -8.12 | -3.68 | -0.91 | -0.18 | -0.05 | 5.35 | 2.26 | 8.13 |
| **2** | -5.43 | -3.48 | -3.54 | 1.30 | -.206 | -3.53 | 5.12 | 3.22 | -2.66 |
| **3** | 1.08 | -5.52 | 1.66 | -7.75 | -4.54 | -0.95 | -1.34 | -5.31 | -9.87 |
| **4** | 0.86 | -3.78 | -4.11 | -5.47 | 0.50 | 3.92 | 4.48 | 3.42 | 5.19 |
| **5** | -2.67 | 0.63 | 7.39 | 6.14 | 5.72 | -4.85 | 7.11 | 2.39 | 9.21 |
| **6** | 4.94 | 3.29 | 2.08 | 3.60 | 1.26 | 4.36 | 7.17 | 4.33 | -0.98 |
| **7** | -2.51 | 2.09 | -2.59 | 5.37 | -4.63 | -3.65 | 5.75 | 3.97 | 6.65 |
| **8** | -2.25 | -2.13 | -6.94 | 7.18 | 1.46 | -6.66 | 0.77 | 0.27 | 2.41 |
| **9** | 5.56 | 2.86 | -2.26 | -7.39 | 1.17 | 6.30 | 0.90 | -0.43 | -8.71 |
| **10** | 1.03 | -3.33 | 4.33 | -7.50 | -6.32 | -0.31 | 3.52 | -0.36 | 6.43 |

(a) 表格 1 为三类样本中的各 10 个样本点，假设每一类均为正态分布。计算每一类样本的均值矢量和协方差矩阵；用不同的颜色在一个图上分别画出这三个类在马氏距离分别为 D=1、2、3 时的图形；为这三个类别的分类设计一个最小马氏距离分类器。

(b) 对测试点：$(1,2,1)^t$，$(5,3,2)^t$，$(0,0,0)^t$，$(1,0,0)^t$，分别计算这些点到各个类的马氏距离，并用（a）中设计的分类器对它们进行分类。

## （3）拓展实验

(c) 对（b）中的测试点，分别计算它们到各个类中心的**欧氏距离**。若使用最小欧氏距离对它们进行分类。将分类结果与(b)中的最小马氏距离分类器的结果进行比较。讨论在什么情况下马氏距离分类器等价于欧氏距离分类器？

(d) 假设三个类别的先验概率相等均为 $P(w_1)=P(w_2)=P(w_3)=1/3$，利用实验 Proj02-01 的方法设计贝叶斯分类器，对（b）中的测试点进行分类，并将其与最小马氏距离分类器的结果进行比较。

(e) 分析实验结果。

## Proj03-01：主分量分析 PCA

### (1) 基本概念介绍

本实验的目的：了解 PCA 主分量分析方法的基本概念，学习和掌握 PCA 主分量分析方法。利用 PCA 分析对数据集合进行特征空间的规整化；利用 PCA 分析对数据集合进行特征空间降维分析。

### (2) 利用 PCA 进行特征空间的规整化

(a) 参考实验 Proj01-01，给定均值矢量和协方差矩阵如下：

$$\mu = \begin{pmatrix} 5 \\ 7 \end{pmatrix}, \qquad \Sigma = \begin{pmatrix} 9 & 2.4 \\ 2.4 & 1 \end{pmatrix}$$

生成 N=100 个高斯分布的二维样本矢量。绘出该样本集合 X 的二维散点图。（可以使用 Matlab 中的函数 plot 画二维散点图）

(b) 利用课本 3.8 节公式，计算上述样本集合 X 的均值向量 m 和散布矩阵 S；使用 matlab 中的函数 eig()，计算 X 的散布矩阵 S 的特征值和特征向量。

(c) 假设数据集合 X 的均值向量为 m，散布矩阵 S 的特征向量矩阵为 $V = [\, e_1 \; e_2 ]^T$，将集合 X 中的每个向量 x 变换为向量 y，生成集合 Y，$\mathbf{y} \in \mathbf{Y}$

$$y = V(x-m)$$

绘出集合 Y 的二维散点图。

(d) 改变样本数量，分别令 N=10、100、1000、10000、100000，重复实验(a)(b)(c)。比较分析集合 X 和 Y 的散点图，说明集合 X 和集合 Y 的关系。并说明该试验中 PCA 方法的意义。

注意：在上述实验中，特别注意 Matlab 的矩阵乘法和向量运算的正确使用方法。

### (3) 利用PCA进行特征空间降维

(e)参考实验Proj01-01，给定均值矢量和协方差矩阵如下：

$$\mu=\begin{pmatrix}10\\15\\15\end{pmatrix},\qquad \Sigma=\begin{pmatrix}90&2.5&1.2\\2.5&35&0.2\\1.2&0.2&0.02\end{pmatrix}$$

生成N=20个高斯分布的三维样本矢量。绘出该样本集合X的三维散点图。(可以使用Matlab中的函数plot3画三维散点图)

(f)利用课本3.8节公式，计算上述样本集合X的均值向量m和散布矩阵S；计算X的散布矩阵S的特征值$\lambda_1$、$\lambda_2$、$\lambda_2$及对应的特征向量$e_1$、$e_2$、$e_3$，其中，$\lambda_1>\lambda_2>\lambda_2$，令特征向量矩阵为V =$[e_1\ e_2\ e_3]^T$。对每一个x计算分量$y_1$、$y_2$，

$$y_1 = e_1^T(x-m),\quad y_2 = e_2^T(x-m)$$

并生成二维向量y=$[y_1\ y_2]^T$,所有的y向量生成了二维向量集合Y，$\mathbf{y}\in\mathbf{Y}$。绘出集合Y的二维散点图。

(g)计算特征向量矩阵V的逆矩阵$V^{-1}=[v_1\ v_2\ v_3]$，其中，$v_1$、$v_2$、$v_3$均是三维列向量。取出逆矩阵$V^{-1}$的前两列组成一个新的3行2列的矩阵W=$[v_1\ v_2]$。对(h)中计算出的每一个二维y向量，用下式计算出一个新的三维z向量

$$z = W\ y + m,$$

所有的z向量生成了三维向量集合Z，$\mathbf{z}\in\mathbf{Z}$。在一幅图中用不同的颜色分别绘出集合Z 和(e)中的集合X的三维数据散点图。

(h)对集合Z中的每一个向量z，及与其对应的集合X中的向量x，计算它们的误差平方值$|x-z|^2$，并计算所有的这些误差平方之和；计算它们的均方误差（误差平方和的平均值）。

(i)改变样本数量，分别令N=10、20、50、100、1000，重复实验(e)(f)(g)(h)。

(j)结合课本中PCA降维的基本思想，以及上面实验中得到的图形和数据，进行深入的比较分析。

## Proj03-02：Fisher 线性判别分析 FDA

### (1) 基本概念介绍

本实验的目的是学习和掌握Fisher线性判别方法FDA。比较分析使用FDA方法寻找的最优投影方向与非最优方向的差异。用FDA方法首先将高维分布的数据降维为低维空间的数据，然后在低维空间中设计分类器进行分类。

两类情况的FDA算法可参见课本3.8.2节内容。多类情况的MDA算法，可参见课本3.8.3节内容。本实验的数据见表格2

表格 2

| 样本 | ω1 | | | ω2 | | | ω3 | | |
|---|---|---|---|---|---|---|---|---|---|
| | $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ |
| **1** | 0.42 | -0.087 | 0.58 | -0.4 | 0.58 | 0.089 | 0.83 | 1.6 | -0.014 |
| **2** | -0.2 | -3.3 | -3.4 | -0.31 | 0.27 | -0.04 | 1.1 | 16. | 0.48 |
| **3** | 1.3 | -0.32 | 1.7 | 0.38 | 0.055 | -0.035 | -0.44 | -0.41 | 0.32 |
| **4** | 0.39 | 0.71 | 0.23 | -0.15 | 0.53 | 0.011 | 0.047 | -0.45 | 1.4 |
| **5** | -1.6 | -5.3 | -0.15 | -0.35 | 0.47 | 0.034 | 0.28 | 0.35 | 3.1 |
| **6** | -0.029 | 0.89 | -4.7 | 0.17 | 0.69 | 0.1 | -0.39 | -0.48 | 0.11 |
| **7** | -0.23 | 1.9 | 2.2 | -0.011 | 0.55 | -0.18 | 0.34 | -0.079 | 0.14 |
| **8** | 0.27 | -0.3 | -0.87 | -0.27 | 0.61 | 0.12 | -0.3 | -0.22 | 2.2 |
| **9** | -1.9 | 0.76 | -2.1 | -0.065 | 0.49 | 0.0012 | 1.1 | 1.2 | -0.46 |
| **10** | 0.87 | -1.0 | -2.6 | -0.12 | 0.054 | -0.063 | 0.18 | -0.11 | -0.49 |

### (2) 基本实验：FDA 分析

(a) 编写一个实现课本3.8.2节的公式（106）的用FDA对三维空间数据求最优方向矢量 **w** 的通用函数。注意，方向矢量 **w** 是一个单位矢量。

(b) 用(a)中的函数对表格2中的类别ω2和ω3，计算最优方向矢量 **w**。

(c) 用课本3.8.2节的公式（91）计算表格2中的类别ω2和ω3的所有数据点在矢量方向 **w** 上的投影。注意，投影后得到ω2和ω3的两个一维数据集合。

(d) 在一幅图中用不同的颜色分别画出表2中类别ω2和ω3数据的三维散点图，同时画出表示方向矢量 **w** 的直线，并且标记出投影后的点在直线上的位置。

(e) 对(c)中投影后得到的两个一维数据集合，假设它们都满足一维高斯分布，分别计算出它们的均值和方差。并假设它们的先验概率P(ω2)=P(ω3)=0.5，设计一个方向 **w** 上的一维贝叶斯分类器。

(f) 用(e)中得到的分类器对 $\omega_2$ 和 $\omega_3$ 的所有数据点进行分类，并计算分类器的训练误差，即错分点的个数。

(g) 比较实验，假设 $\mathbf{v}=(1.0,2.0,-1.5)^T$，令非最优方向矢量 $\mathbf{w}=\mathbf{v}/\|\mathbf{v}\|$，重复(c)(d)(e)(f)步骤，计算在这个非最优子空间中，分类器的训练误差。

(h) 对实验结果进行分析。

### (3) 拓展实验：MDA 分析

(i) 编写一个求解课本 3.8.3 节的公式（126），计算用 MDA 对三维空间的三个类的数据求最优二维投影子空间 $\mathbf{W}=[\mathbf{w}_1, \mathbf{w}_2]$的通用函数。其中，子空间 $\mathbf{W}$ 中的两个方向基矢量 $\mathbf{w}_1$、$\mathbf{w}_2$ 分别是公式（126）中矩阵 $\mathbf{S}_W^{-1}\mathbf{S}_B$ 的最大特征值和次大特征值对应的特征矢量，方向基矢量 $\mathbf{w}_1$、$\mathbf{w}_2$ 均为单位矢量。

(j) 用(a)中的函数对表格 2 中的三个类别 $\omega_1$、$\omega_2$ 和 $\omega_3$ 的数据，计算最优二维投影子空间 $\mathbf{W}=[\mathbf{w}_1, \mathbf{w}_2]$。

(k) 用课本 3.8.3 节的公式（117）或公式（118），计算表格 2 中的三个类别 $\omega_1$、$\omega_2$ 和 $\omega_3$ 的所有数据点在二维子空间 $\mathbf{W}$ 上的投影。注意，投影后分别得到数据集 $\omega_1$、$\omega_2$ 和 $\omega_3$ 的一个二维数据集合。

(l) 在一幅图中用不同的颜色画出表 2 中三个类别数据的三维散点图。

(m) 在一幅图中用不同的颜色画出(k)中计算出的三个类别数据的二维散点图。

(n) 对(k)中投影后得到的三类二维数据集合，假设它们都满足高斯分布，分别计算出它们的均值向量和协方差矩阵。若它们的先验概率 $P(\omega_1)=P(\omega_2)=P(\omega_3)=1/3$，设计子空间 $\mathbf{W}$ 上的贝叶斯分类器。

(o) 用(n)中得到的分类器对 $\omega_1$、$\omega_2$ 和 $\omega_3$ 的所有数据点进行分类，并计算分类器的训练误差，即错分点的个数。

(p) 比较实验，假设 $\mathbf{v}_1=(1.0,2.0,-1.5)^T$，$\mathbf{v}_2=(-1.0,0.5,-1.0)^T$，令非最优子空间基矢量 $\mathbf{w}_i=\mathbf{v}_i/\|\mathbf{v}_i\|$，i=1,2。重复(k)(l)(m)(n)步骤，计算在这个非最优子空间 $\mathbf{W}=[\mathbf{w}_1, \mathbf{w}_2]$中，设计的贝叶斯分类器的训练误差。

(q) 对实验结果进行分析。

## Proj04-01：Parzen 窗估计、k 近邻估计

本实验的目的是学习和掌握两种非参数估计方法——Parzen 窗估计、k 最近邻估计。实验使用表格 3 中的数据。假设先验概率 P(ω1)= P(ω2)= P(ω3)= 1/3。

**表格 3**

| 样本 | $w_1$ | | | $w_2$ | | | $w_3$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ |
| **1** | 0.28 | 1.31 | -6.2 | 0.011 | 1.03 | -0.21 | 1.36 | 2.17 | 0.14 |
| **2** | 0.07 | 0.58 | -0.78 | 1.27 | 1.28 | 0.08 | 1.41 | 1.45 | -0.38 |
| **3** | 1.54 | 2.01 | -1.63 | 0.13 | 3.12 | 0.16 | 1.22 | 0.99 | 0.69 |
| **4** | -0.44 | 1.18 | -4.32 | -0.21 | 1.23 | -0.11 | 2.46 | 2.19 | 1.31 |
| **5** | -0.81 | 0.21 | 5.73 | -2.18 | 1.39 | -0.19 | 0.68 | 0.79 | 0.87 |
| **6** | 1.52 | 3.16 | 2.77 | 0.34 | 1.96 | -0.16 | 2.51 | 3.22 | 1.35 |
| **7** | 2.20 | 2.42 | -0.19 | -1.38 | 0.94 | 0.45 | 0.60 | 2.44 | 0.92 |
| **8** | 0.91 | 1.94 | 6.21 | -0.12 | 0.82 | 0.17 | 0.64 | 0.13 | 0.97 |
| **9** | 0.65 | 1.93 | 4.38 | -1.44 | 2.31 | 0.14 | 0.85 | 0.58 | 0.99 |
| **10** | -0.26 | 0.82 | -0.96 | 0.26 | 1.94 | 0.08 | 0.66 | 0.51 | 0.88 |

(1) Parzen 窗估计

设计基本 Parzen 窗估计分类器，以及设计概率神经网络（PNN）分类器。本实验中采用的窗函数为窗口宽度值 $h$ 相同的球状高斯函数，：

$$\varphi\left(\frac{\mathbf{x}-\mathbf{x}_i}{h}\right)=\mathrm{e}^{\frac{-(\mathbf{x}-\mathbf{x}_i)^T(\mathbf{x}-\mathbf{x}_i)}{2h^2}}$$

类条件概率密度估计公式采用下面的公式

$$p_n(\mathbf{x}|\omega)=\frac{1}{n}\sum_{i=1}^{n}\varphi\left(\frac{\mathbf{x}-\mathbf{x}_i}{h}\right)$$

(a) 编写程序，设计一个采用上面的 Parzen 窗估计方法计算三维数据类条件概率密度的程序，并设计一个对三个类分类的贝叶斯分类器。

(b) 编写程序，对于三维数据，设计一个实现课本 142 页的算法 1，以及课本 143 页的算法 2 的可以对三个类分类的概率神经网络 PNN 分类器。其中，在 143 页算法 2 中选择方差参数 $\sigma=h$ 。

(c) 用表格 3 中的三个类的三维数据作为分类器的训练数据，同时，令 $h=1$，分别对 (a) 和 (b) 中设计的两个分类器进行训练。然后用训练后的分类器对样本点 $(0.5，1.0，0.0)^T$，$(0.31，1.51，-0.5)^T$，$(-0.3，0.44，-0.1)^T$ 进行分类。

(d)令 $h=0.1$，重复（a）（b）（c）。

(e)分析实验结果。并分析（a）、（b）中的两种算法异同。

（2）k 最近邻估计

学习和掌握非参数估计——k 近邻概率密度估计方法。对表格 3 中的数据进行 k 最近邻概率密度估计，并进行分类器设计。

其中，$d$ 维空间的距离度量 $L_p$ 范数（ $p>0$）为

$$L_p(\mathbf{x},\mathbf{y})=\left(\sum_{i=1}^{d}|x_i-y_i|^p\right)^{1/p}$$

(a)编写程序，采用欧氏距离度量，对具有 $n$ 个训练样本点的一维数据，实现对任给测试点 x 的 k 近邻概率估计。用编写的程序对表格 3 中的类别 $w_3$ 的特征 $x_1$，分别画出当 k=3、5 时用程序计算出的概率密度估计的结果图。

(b)编写程序，采用欧氏距离度量，对具有 $n$ 个训练样本点的二维数据，实现对任给测试点 x 的 k 近邻概率估计。对表格 3 中的类别 $w_2$ 的特征 $(x_1,x_2)^t$，分别画出当 k=3、5 时用程序计算出的概率密度估计的结果图。提示：可以在数据的扩大的动态范围内生成数据网格，计算网格上的概率估计值，然后画出网格三维图。

(c)编写程序，采用欧氏距离度量，对已标记的具有三个类（三个类的数据个数分别为 $n_1$、$n_2$、$n_3$）的三维训练数据，设计一个如课本 4.5.4 节 k 近邻规则所述的 k 近邻后验概率密度估计分类器，实现对任给测试点 x 的分类。利用表格 3 中的 3 个类别的数据作为训练数据，在分别取 k=3、5 时，对下列测试点进行分类： $(-0.41，0.82，0.88)^T$，$(0.14，0.72，4.1)^T$，$(-0.81，0.61，-0.38)^T$。

(d)分别使用 $L_1(\mathbf{x},\mathbf{y})=\sum_{i=1}^{d}|x_i-y_i|$，$L_\infty(\mathbf{x},\mathbf{y})=\max(|x_i-y_i|,i=1,2,...,d)$ 两个距离度量重新进行(c)中的实验。

(e)分析实验结果。

(3) 试比较分析一下 Parzen 窗估计和 k 最近邻估计各有什么特点。

## Proj05-01：感知器算法

本实验的目的是学习和掌握两种感知器算法。实验使用表格 4 中的数据。

表格 4

| 样本 | $w_1$ | | $w_2$ | | $w_3$ | | $w_4$ | |
|---|---|---|---|---|---|---|---|---|
| | $x_1$ | $x_2$ | $x_1$ | $x_2$ | $x_1$ | $x_2$ | $x_1$ | $x_2$ |
| **1** | 0.1 | 1.1 | 7.1 | 4.2 | -3.0 | -2.9 | -2.0 | -8.4 |
| **2** | 6.8 | 7.1 | -1.4 | -4.3 | 0.54 | 8.7 | -8.9 | 0.2 |
| **3** | -3.5 | -4.1 | 4.5 | 0.0 | 2.9 | 2.1 | -4.2 | -7.7 |
| **4** | 2.0 | 2.7 | 6.3 | 1.6 | -0.1 | 5.2 | -8.5 | -3.2 |
| **5** | 4.1 | 2.8 | 4.2 | 1.9 | -4.0 | 2.2 | -6.7 | -4.0 |
| **6** | 3.1 | 5.0 | 1.4 | -3.2 | -1.3 | 3.7 | -0.5 | -9.2 |
| **7** | -0.8 | -1.3 | 2.4 | -4.0 | -3.4 | 6.2 | -5.3 | -6.7 |
| **8** | 0.9 | 1.2 | 2.5 | -6.1 | -4.1 | 3.4 | -8.7 | -6.4 |
| **9** | 5.0 | 6.4 | 8.4 | 3.7 | -5.1 | 1.6 | -7.1 | -9.7 |
| **10** | 3.9 | 4.0 | 4.1 | -2.2 | 1.9 | 5.1 | -8.0 | -6.3 |

(1) 基本概念介绍

考虑一个由 $d$ 维向量 $\mathbf{x}$ 组成的样本集合 X，该数据集合由两类已经标记的样本组成。感知器算法通过学习这两类已标记的样本，建立一个线性分类器。

基本感知器的准则函数：

$$J_p(\mathbf{a}) = \sum_{\mathbf{y}\in \mathrm{Y}} \left(-\mathbf{a}^{\mathrm{T}}\mathbf{y}\right)$$

则梯度下降迭代公式为：

$$\mathbf{a}(k+1) = \mathbf{a}(k) + \eta(k) \sum_{\mathbf{y}\in \mathrm{Y}_k} \mathbf{y}$$

其中，Y 为被 a 错分样本集合，$\mathrm{Y}_k$ 为被 a(k)错分样本集合。

裕量松弛感知器的准则函数：

$$J_r(\mathbf{a}) = \frac{1}{2}\sum_{\mathbf{y}\in \mathrm{Y}} \frac{\left(\mathbf{a}^{\mathrm{T}}\mathbf{y} - b\right)^2}{\|\mathbf{y}\|^2}$$

则梯度下降迭代公式为：

$$\mathbf{a}(k+1)=\mathbf{a}(k)+\eta(k)\sum_{\mathbf{y}\in \mathrm{Y}_k}\frac{b-\mathbf{a}(k)^{\mathrm{T}}\mathbf{y}}{\|\mathbf{y}\|^2}\mathbf{y}$$

其中，Y 为被 a 错分样本集合，$Y_k$ 为被 a(k)错分样本集合。

**(2) 批处理感知器算法（课本 186 页算法 3）**

(a) 编写程序实现本章 186 页算法 3 的批处理感知器算法。

(b) 将你编写的程序应用在表格 4 中 $w_1$ 和 $w_2$ 的训练数据上。从 $\mathbf{a}=\mathbf{0}$ 向量开始，记下每一次迭代时的 $J_p(a)$，记下收敛时的迭代步数。绘出 $J_p(a)$ 的值随迭代次数变化的曲线。

(c) 将程序应用在 $w_2$ 和 $w_3$ 的训练数据上，重复执行步骤(b)。

(d) 分析实验(b)、(c)的结果。

(e) 改变学习率 η。然后重复上述(b)、(c)、(d)的实验。

**(3) 批处理裕量松弛算法（课本 193 页算法 8）**

(a)编写程序实现本章 193 页的算法 8 的批处理裕量松弛感知器算法

(b)对在表格 4 中 $w_1$ 和 $w_3$ 的训练数据进行批处理松弛算法。设间隔 b=0.1，初始权向量为零向量 a(1)=0，画出准则函数对于训练回合数的函数曲线。

(c)设间隔 b=0.5，初始权向量为零向量 a(1)=0，重复（b）。画出准则函数对于训练回合数的函数曲线。

(d)使用具体的实验数据解释(b)、(c)在收敛率上的不同。

(e)改变学习率 η。然后重复上述(b)、(c)、(d)的实验。

**(4) 分析比较在实验（2）和（3）中的算法各有什么特点。**

## Proj06-01： BP 神经网络

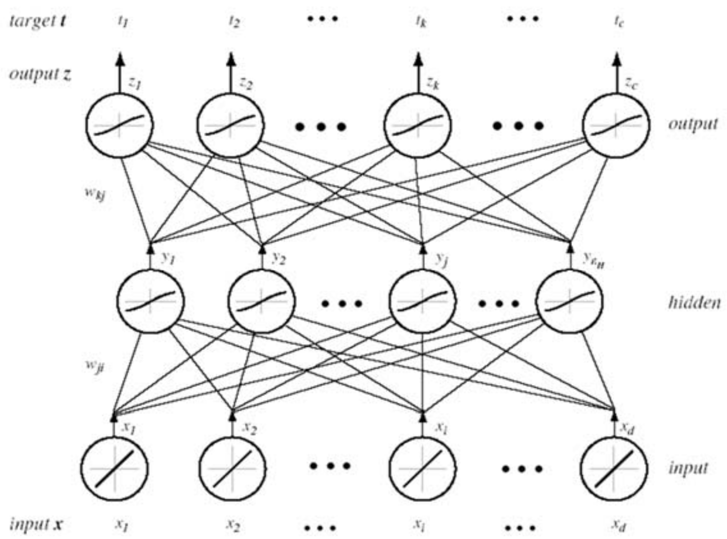
本实验的目的是学习和掌握 BP 神经网络原理、学习算法，及其应用。本实验使用表格 5 中的数据：

表格 5

| 样本 | $\omega_1$ | | | $\omega_2$ | | | $\omega_3$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ | $x_1$ | $x_2$ | $x_3$ |
| **1** | 1.58 | 2.32 | -5.8 | 0.21 | 0.03 | -2.21 | -1.54 | 1.17 | 0.64 |
| **2** | 0.67 | 1.58 | -4.78 | 0.37 | 0.28 | -1.8 | 5.41 | 3.45 | -1.33 |
| **3** | 1.04 | 1.01 | -3.63 | 0.18 | 1.22 | 0.16 | 1.55 | 0.99 | 2.69 |
| **4** | -1.49 | 2.18 | -.39 | -0.24 | 0.93 | -1.01 | 1.86 | 3.19 | 1.51 |
| **5** | -0.41 | 1.21 | -4.73 | -1.18 | 0.39 | -0.39 | 1.68 | 1.79 | -0.87 |
| **6** | 1.39 | 3.16 | 2.87 | 0.74 | 0.96 | -1.16 | 3.51 | -0.22 | -1.39 |
| **7** | 1.20 | 1.40 | -1.89 | -0.38 | 1.94 | -0.48 | 1.40 | -0.44 | 0.92 |
| **8** | -0.92 | 1.44 | -3.22 | 0.02 | 0.72 | -0.17 | 0.44 | 0.83 | 1.97 |
| **9** | 0.45 | 1.33 | -4.38 | 0.44 | 1.31 | -0.14 | 0.25 | 0.68 | -0.99 |
| **10** | -0.76 | 0.84 | -1.96 | 0.46 | 1.49 | 0.68 | -0.66 | -0.45 | 0.08 |

(1) 基本概念

考虑一个 $d$-$n_H$-$c$ 的三层 BP 网络，网络有 $d$ 个输入层神经元，有 $n_H$ 个中间层神经元，有 $c$ 个输出层神经元。并规定输入层为线性神经单元。




定义网络的误差函数：$J(\mathbf{w}) \equiv \frac{1}{2}\sum_{k=1}^{c}(\mathbf{t}_k - \mathbf{z}_k)^2 = \frac{1}{2}\|\mathbf{t}-\mathbf{z}\|^2$

神经元函数：$f(y) = \frac{1}{1+e^{-y}}$，　神经元函数的导数：$f'(y) = f(y)(1-f(y))$

中间层神经元输出：$y_j = f\left(\sum_{i=1}^{d} w_{ji} x_i + w_{j0}\right)$，$j = 1,2,\cdots,n_H$

输出层神经元输出：$z_k = f\left(\sum_{j=1}^{n_H} w_{kj} y_j + w_{k0}\right) = f\left(\sum_{j=1}^{n_H} w_{kj} f\left(\sum_{i=1}^{d} w_{ji} x_i + w_{j0}\right) + w_{k0}\right)$

权值更新：$w(m+1) = w(m) + \Delta w(m)$，$\Delta w = -\eta \frac{\partial J}{\partial w}$

其中：$\Delta w_{kj} = -\eta \frac{\partial J}{\partial w_{kj}} = \eta (t_k - z_k) \cdot f(net_k)(1 - f(net_k)) \cdot y_j$

$$\Delta w_{ji} = -\eta \frac{\partial J}{\partial w_{ji}} = \eta \sum_{k=1}^{c} \left[(t_k - z_k) \cdot f'(net_k) \cdot w_{kj}\right] \cdot f'(net_j) \cdot x_i$$

**(2) 简单三层 BP 神经网络**

(a)构造一个 3-3-1 型的三层 BP 神经网络，参数包括：输入层、中间层、输出层神经元向量，以及输入层到中间层的权值矩阵，中间层到输出层的权值矩阵，中间层神经元的偏置向量，输出层神经元的偏置向量等。

(b)编写函数，实现课本 233 页中式(7)所述的 BP 网络的前馈输出。为了便于后续的学习算法的实现，该函数的输出变量可以包含网络中间节点的输出。

(c)编写函数，实现课本 237 页中式(12)、(17)、(21)所述的 BP 网络的权值修正。

(d)编写函数，实现课本 236 页中式(9)所述的 BP 网络的训练误差。

(e)编写函数，实现课本 239 页中算法 1 所述的 BP 网络的训练算法。

(f)用表格 5 中的$\omega_1$和$\omega_2$类的数据进行训练，学习率$\eta = 0.1$。在$-1 \le w \le +1$范围内随机初始化所有权值。绘出训练误差对训练回合数变化的学习曲线。其中，一个回合是指用所有数据训练一次。用 t=0、1 分别表示两个类教师信号。

(g)初始化所有权值为$w = 0.5$。重复（f）。

(h)分析实验结果。

**(3) 拓展实验：可用于三个类分类的三层 BP 神经网络**

(i)构造一个 3-4-3 型的三层 BP 神经网络。用表格 5 中的$\omega_1$、$\omega_2$和$\omega_3$类的数据进行训练。用 $t_1$=(1,0,0)，$t_2$=(0,1,0)，$t_3$=(0,0,1),分别表示三个类的输出层教师信号。使用不同的权值初始化方法进行训练，绘出训练误差对训练回合数变化的学习曲线。

(j)分析实验结果。