**1.**

已知训练集

设隐层两个突触的权值、偏置值、阈值分别为(, , , )、(, , , )

设输出层突触的权值、偏置值、阈值为(, , , )

由于需要得到对于四个输入的无差错感知机参数，样本数少且重要，因而不妨采用较大的学习率，故设学习率

由于反向传播算法对于任意参数的更新估计式为



因而对于偏置值，有



由于



记，故有



同理可得



对其余参数，同理可得









由于



故不妨设初始值







以下演示一组计算过程，对于该感知机，有



代入第一组训练样本

由





得









由





得





由



得



由



得





综上，又由









得











故更新后的参数为













如此循环代入四个样本，具体程序代码见1.py，最终得到的参数为













此时若代入四组数据，得到的神经网络输出值为

时，

时，

时，

时，

使用阈值函数



即可得到符合预期的输出。

**2.**

**(1).**

**指出Fisher线性判别中，w的比例因子对Fisher判别结果无影响的原因：**

线性判别的思想在于，让同类样例投影点的协方差尽可能小，即令类内散度矩阵尽可能小。同时，让不同类中心之间的距离尽可能大，即类间散度矩阵尽可能大。

但该问题难以直接求解，因而引入了正交投影矩阵，将该问题降维，借助广义瑞利商的概念，将问题转化为最大化

因而线性判别的结果，也即对于最大化的问题而言，作为一个正交投影矩阵，结果只与正交投影矩阵的投影方向的因子有关，而与正交投影矩阵的比例因子无关。

**分析J(w)可用Lagrange乘子法求解的条件：**

将等式约束代入，得



假设使取得极值的解为，则由KKT条件可知



由于可导，因而只需确定对可导。

若样本的维数为，则是的矩阵，矩阵对矩阵求导后有、。从几何意义上而言，也即对原有的样本进行一次降维的投影变换。

故求解条件为——是一个投影矩阵。

**(2).**

随机变量的分布律为



故似然函数为



而





故



**3.**

鸢尾花数据集中域包含四个属性（不妨命名为），在域中将其划分为了三种花，按照信息熵的定义，有



由于数据集为连续值，为了应用最大信息增益算法，假定每个属性值只产生2个分支结点，即，分别代表大于、小于划分点值的两种情况



因此，首先对四个属性值分别进行二分离散化，方法为：

对于连续属性在中的个不同取值，先将这些值从小到大排序，记为，基于划分点可将分为子集和，其中表示在属性上不大于的样本，反之。由此，我们可考察包含个元素的候选划分点集合



由此，对于属性的信息增益，可取为



以下依次为程序运行后，基于训练集的决策树、经过测试集预剪枝后的决策树：

 