

6.034 测验 2

2000 年 秋

姓名:	
电子邮件:	
助教:	
指导教师:	

题号	得分:
题1	
题2	
题3	
总分	

题 1：选择题（30 分）

这个题放在最前面是因为其它的题都会比这花 3 到 4 倍的时间。给最符合题意的唯一选项划圈。多选不得分。

逐步加深法，也称为反复加深法，在博弈中应用出色是因为：

$\alpha - \beta$ 方法在给出的博弈树可以前进 2 倍深度。

分支因子逐层变化

给定深度的博弈树中几乎所有节点都在最深的一层

给定深度的博弈树中几乎没有任何一个节点在最深的一层。

以上皆是

以上皆不是

$\alpha - \beta$ 法：

速度是最小最大法的的两倍

速度比最小最大法慢

和最小最大法矛盾

和逐步加深法相矛盾

以上皆是

以上皆不是

拓扑分类算法是为了：

改善运行速度

确保连接点原则决定优先权

处理继承树中的循环

实现排序原则

以上皆是

以上皆不是

帧用来：

理解句子

理解故事

理解隐喻

通过继承进行默认的推理

以上皆是

以上皆不是

空间转换表示法的一个关键优点是：

包含了主题帧和原始动作帧

能够转换成关系数据库里的纪录

使状态可以描述

使隐喻更好理解

以上皆是

以上皆不是

主题帧表示法的一个关键优点是：

用几个基本的动作表达所有的动作。

以故事的方式描述

把焦点放在变量的变化上

抓住事物的因果关系

以上皆是

以上皆不是

语义转换树语法的关键优点是：

减少了需要理解的词汇数量

用递归代替循环，简化了语法结构

采用 Rete 算法

采用空间转换表示法

以上皆是

以上皆不是

自然语言数据库接口起作用的原因是因为处理名词短语总是变成：

并运算

选择运算

投影运算

分类运算

以上皆是

以上皆不是

遗传算法中交换的目的是：

增加多样性

为自然突变建立模型

改变选择的可能性

减少遗传型的数目

以上皆是

以上皆不是

在神经网络中：

生物神经网络能精确模拟

S 型阈值用来避免过拟合

过少的训练周期将导致过拟合

每个训练周期中反向传播的计算量和 n^2 成正比，其中 n 是节点数

以上皆是

以上皆不是

题 2：最近邻，决策树和支持向量（30 分）

议会决定问每个选举人几个关键的问题以预测选举人是怎么投票的。显然这将为每个人

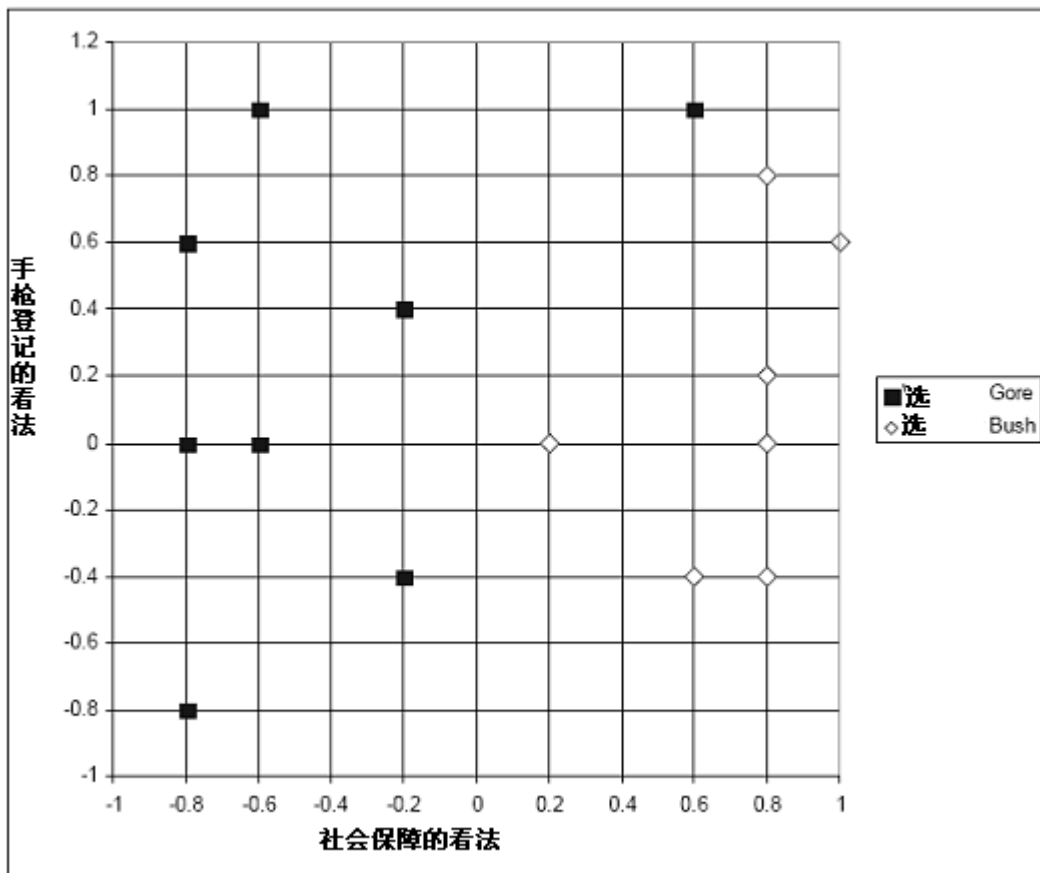
节省检查选举票计算选举结果时间和精力。

他们决定从下面两个问题开始：

- a) 在-1 到+1 的范围内，-1 表示强烈反对，+1 表示强烈同意，你对个人社会保障有何看法？
- b) 在-1 到+1 范围内，你对手枪登记有何看法？

下面给出的是训练例子，15 个人对这两个问题的看法示在图中标出，实心方框投的是 Gore，空心菱形投的是 Bush，例子显示出了他们是怎样投票选举总统的。

坐标中还有“？”的标记，这表示的是个未决定选谁的选举人。我们把他叫做不确定先生。



A 部分：最近邻法（10 分）

1)最近邻法预测不确定先生会选谁？假设采用标准欧几里德距离为度量标准。（你的答案应该是 Gore 或者 Bush）

2)在上图中，仔细画出精确的最近邻法显示的边界线，以区分 Gore 的样本空间和 Bush 的样本空间。在分析中不包含“？”。

3)最近邻法预测不确定先生的投票结果是什么？（运用同样的欧几里德度量标准）

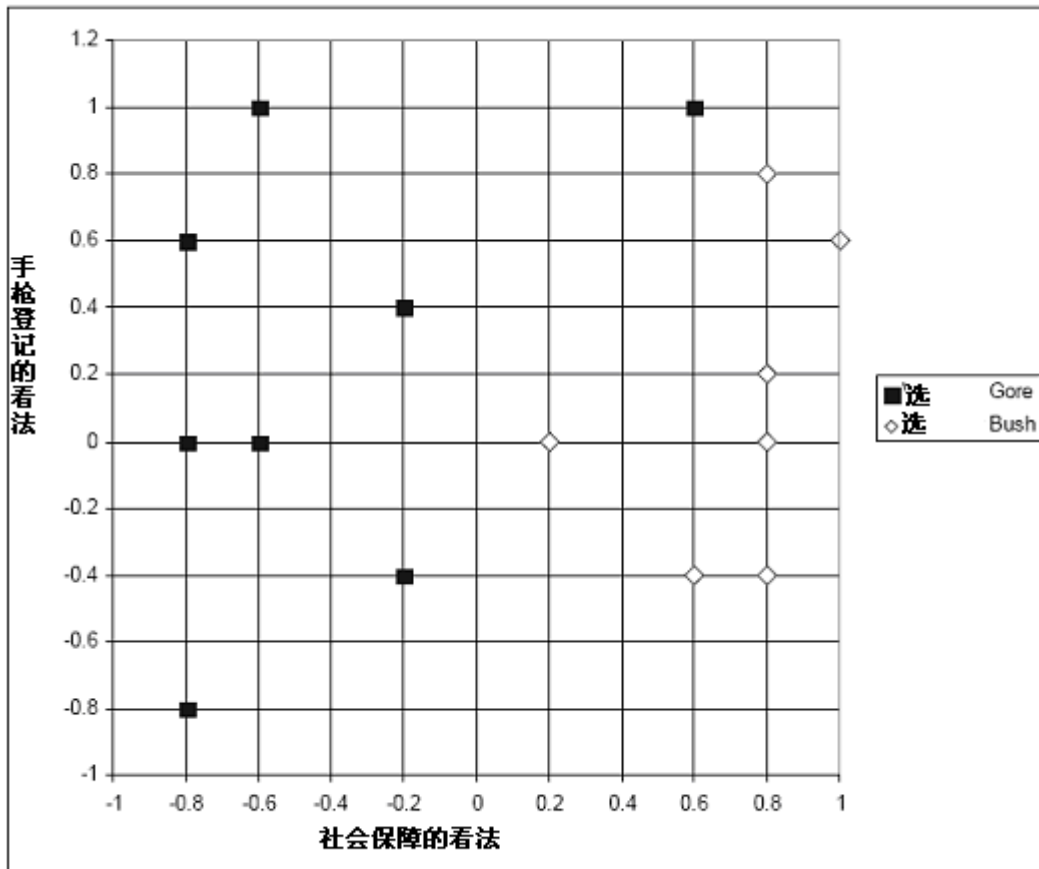
4)另外一个必须询问选举人的问题产生了：“你对降低议会议员的工资有何看法？”这个问题不包括在公开数据中，因为（根据政客的观点）这些数据对做决定没有什么作用。更

深一层次的思考发现实际的问题是所有的结果都聚集在+1 附近。作为顾问你建议：

- a) 政客说的对，这些数据毫无意义
- b) 数据还是有意义的，你只需要把所有值都除以其平均值。
- c) 数据还是有用的，只需要把所有值都减去其平均值，然后再乘以标准偏差。
- d) 数据还是有用的，只需要把所有值都除以其标准偏差。
- e) 数据是有用的，不过上面的方法都不对。

B 部分 判定树 (8 分)

为了你的方便这里再重复同样的数据。



事情似乎进展的很顺利，忽然，Ralph Nader 出现了，他建议最近邻法就是在浪费宝贵的时间和空间，应该用判定树取而代之。

1) 你决定按照最先的试验对社会保障的看法小于零的情况进行试验。但是，你知道，你需要确定试验数据的平均无序度，用来判断其好坏。那么平均无序度是多少啦？

(答案可能包含 \log_2 算子，你没有必要化简表达式)。

2) 试验看起来不错，你决定完成这个决策树。那么画出判定树，便详细说明每个试验。分析中不能含有“？”。

3) 你的决策树预测不确定先生会选哪位总统？在判定树中圈出相应的节点。

C 部分 更多投票人问题 (4 分)

发明因特网技术的 Mr. Gore 称对科学技术略知一二。他说我们对投票人提的问题太少了，要得到相当好的预测能力，我们应该调查至少 100 个问题。你提出 100 个问题，你为此正努力工作，却不认为这 100 个问题会给你关于选举人的预测信息。不过你还是尝试运用最近邻和判定树两种方法。你的初步试验揭示：（勾出最有可能的结果）

- a) 两种技术都很好而且难分伯仲。
- b) 最近邻比判定树要好
- c) 判定树比最近邻好
- d) 两者都不行

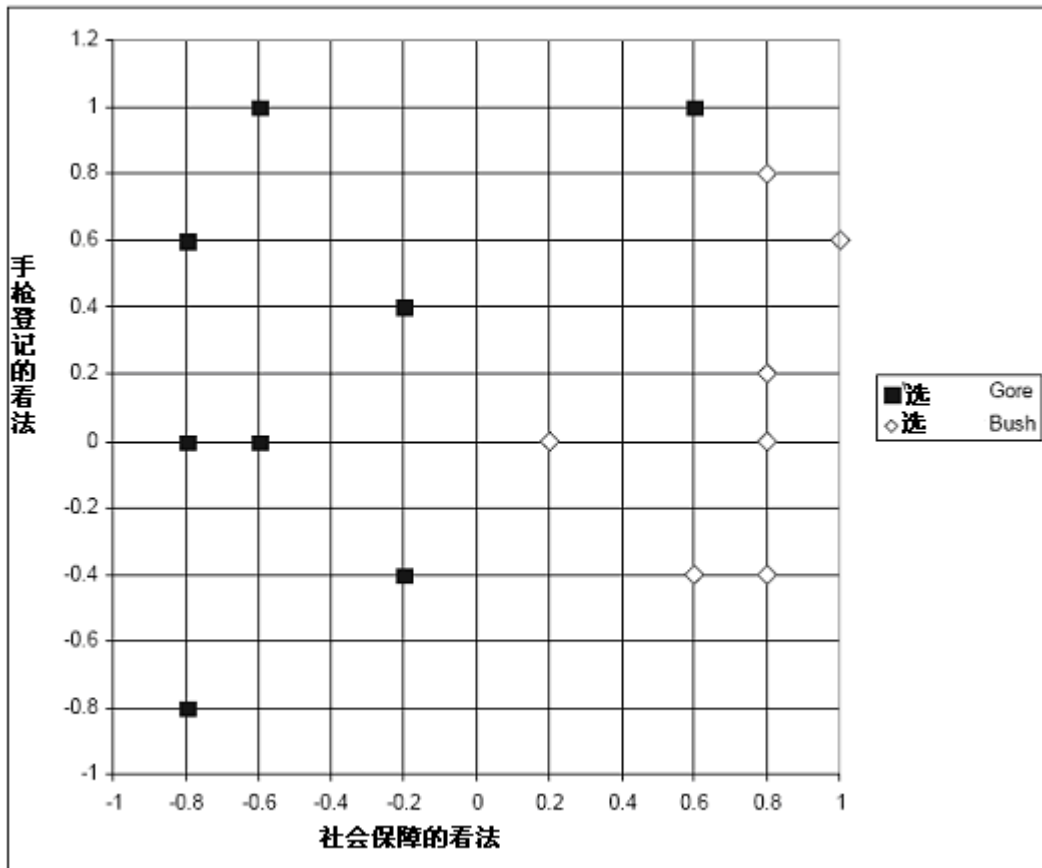
D 部分 支持向量机 （8 分）

一切都将结束的时候你注意到了街上一个很小但是响亮的示威，示威的人们都带有叫做支持向量机的标记。你有点困惑，向同事请教什么是向量机？他解释说你的说法有点欠妥，应该叫做支持向量机。问题又回到你这边，你想通过下面的问题来搞清楚答案。

你决定采用径向基函数开始试验：

$$K(x_1, x_2) = e^{-\frac{\|x_1 - x_2\|^2}{2\delta^2}}$$

你的试验成功了，得到了一些少量的形成支持向量的样本。所有的样本都小于等于-1或大于等于+1。在下面的图中绘制出辨别边界：



2. 圈出下列说法中正确的选项:

有一个 δ 值使所有的 Gore/Bush 样本都将成为支持向量。

有一个 δ 值使 Gore/Bush 样本不能分开。

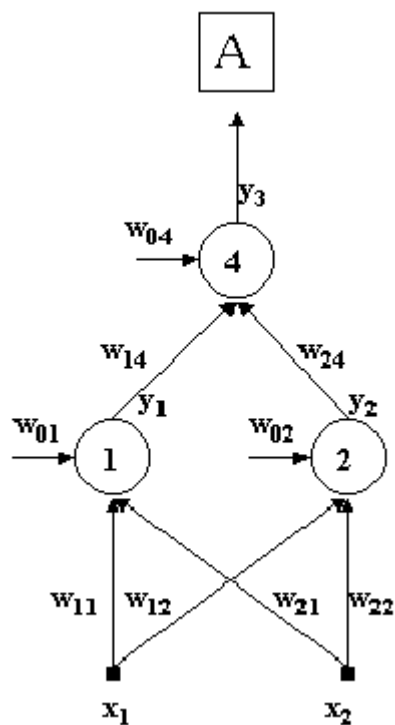
有一个 δ 值会产生过拟合。

如果用 $x_1 \cdot x_2$ 点积代替径向基函数, Gore/Bush 样本将可分开。

题 3: 神经网络 (40 分)

A 部分 (26 分)

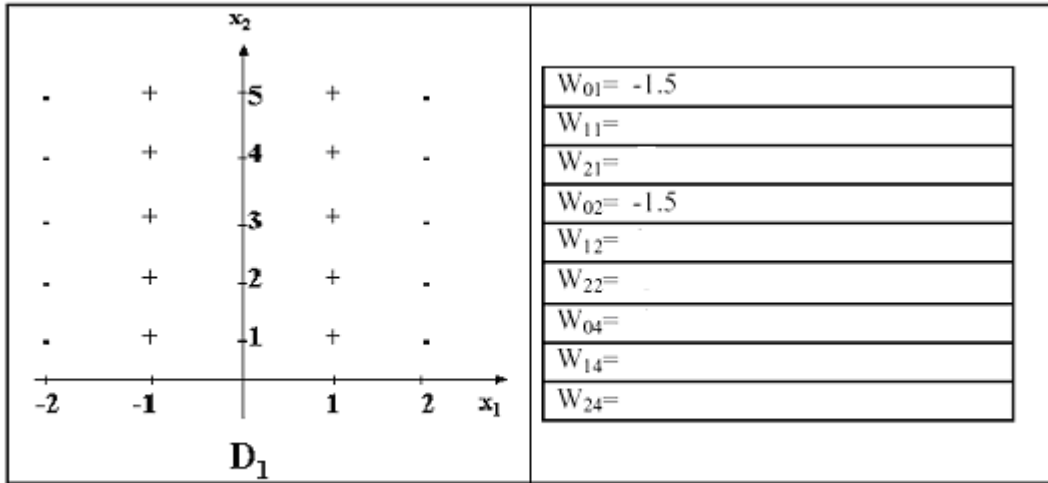
考虑下面的神经网络, 3 个都是感知器而不是 S 型单元, 也就是说, 所有单元的输出 0 或 1。假设所有的输入都有权值, 比如 w_{01} 和 w_{02} 都是 -1。



A.1 (10 分)

补充 A 网络中缺少的权值以把 D1 中的数据正确分类。

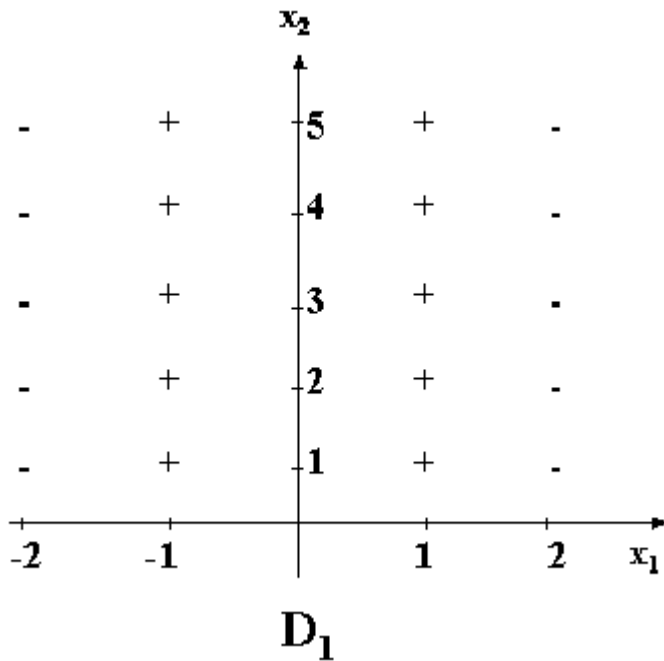
例如标记-的输出为 0, 标记+的输出为 1。



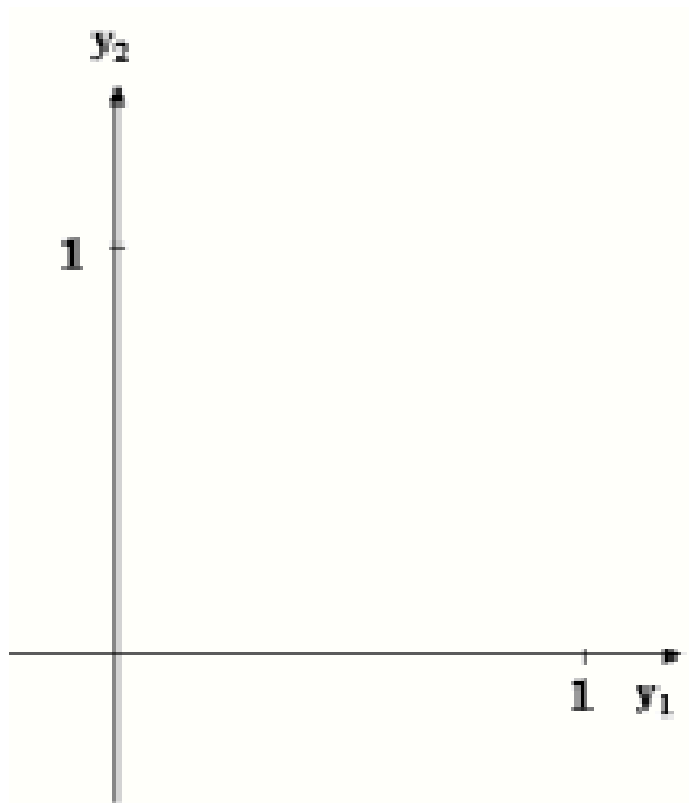
A.2(8分)

根据 A.1 中计算出的权值：

1. 在下面的图中仔细画出 1 与 2 单元的决策边界。对每个边界标上 1 或者 2。

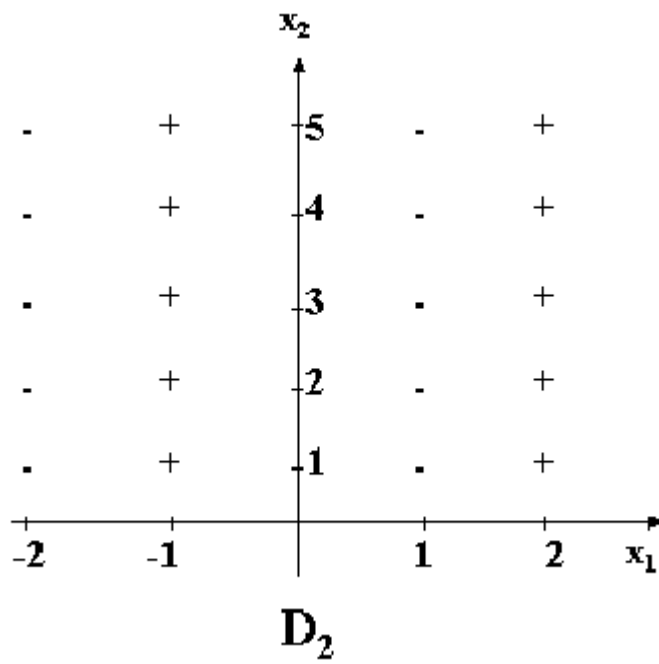


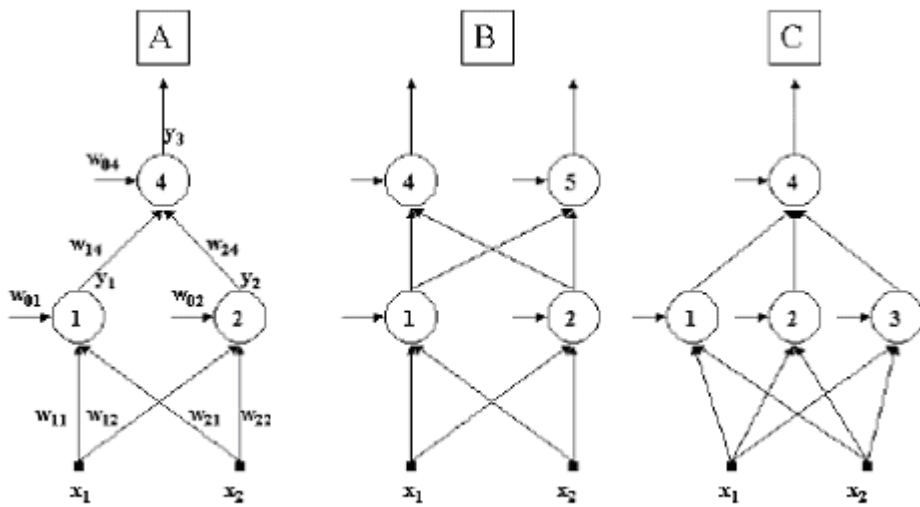
2. 在下图中画出单元 1 和单元 2 的输出组合。用+或者-或者两者皆有标记每种组合，给出 D_1 的数据集。



A.3(8分)

考虑下面的数据集，和 D1 略有不同。

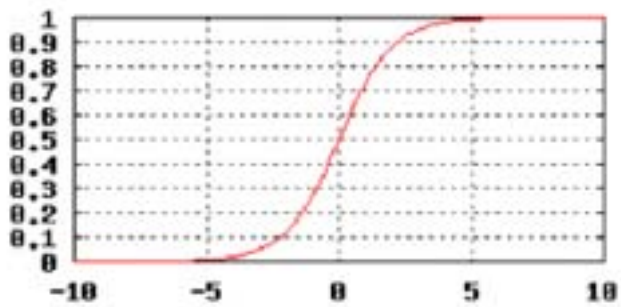
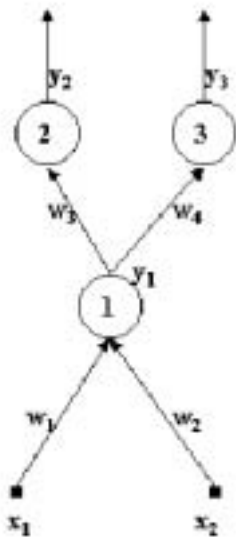




圈出能够正确对这个数据集分类的网络结构。

假设所有结构都采用感知单元。

B 部分： (14 分)



S 型函数值:

S(-5)=0	S(-4)=0.02	S(-3)=0.04	S(-2)=0.12	S(-1)=0.27	S(0)=0.5
S(5)=1	S(4)=0.98	S(3)=0.96	S(2)=0.88	S(1)=0.73	

B.1 (6 分)

假设每个单元都有固定的阈值 0.0, 同时假设输入都为 2, 权值 $w_1=2, w_2=-2, w_3=4, w_4=0$, 那么计算 (大约):

$$y_1 =$$

$$y_2 =$$

$$y_3 =$$

B.2(8分)

现在碰到了两种情形，你打算判断一下权值是增大还是减小。在两种情形中，你将运用 B.1 的网络结构，这个网络用的是 S 型神经元。

情形 1: 假设期望的输入输出对应关系如下:

$$x_1 = 2,$$

$$x_2 = 2$$

时,

$$y_2 = 0,$$

$$y_3 = 1,$$

设学习率为 1, 权值同 B.1.

情形 2: 假设同样的输入希望的输出值为:

$$y_2 = 1,$$

$$y_3 = 0,$$

用增或减的形式正确填写下列表格:

	情形1	情形2
W1		
W2		
W3		
W4		

下页给出了一些有用的信息，你可以将其撕下以供参考。

反向传播

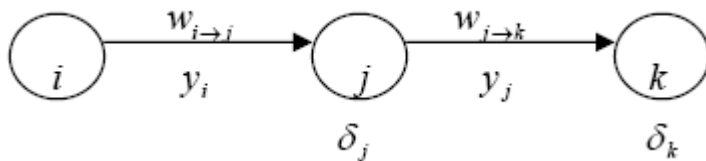
实现神经网络梯度下降的有效方法

下降规则

$$w_{i \rightarrow j} = w_{i \rightarrow j} - r \frac{dE}{dw_{i \rightarrow j}} = w_{i \rightarrow j} - r \delta_j y_i$$

$$\delta_j = \frac{ds(z_j)}{dz_j} \sum_k \delta_k w_{j \rightarrow k}$$

背向传播规则：

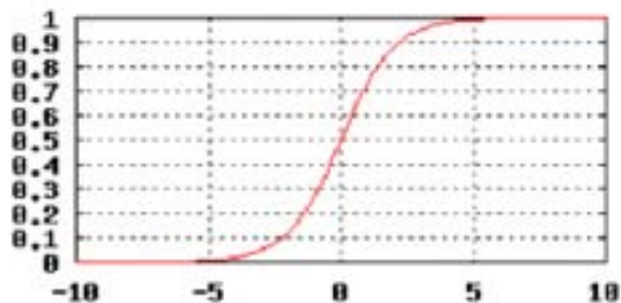
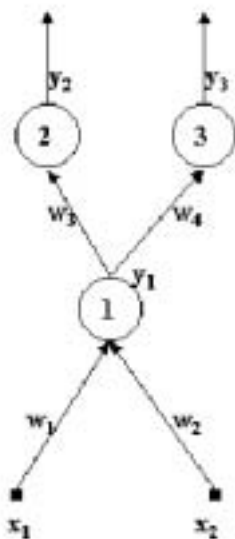


y_i 是 x_i 的输入层

- 1、将权值初始化为比较小的随机值；
- 2、选择一个随机样品的输入特征矢量；
- 3、为每个单元（前向传播）计算所有的输入值（ z_j ）和输出值（ y_j ）

$$\delta_n = \frac{ds(z_n)}{dz_n} (y_n - y_n^*) = y_n (1 - y_n) (y_n - y_n^*)$$

- 4、为输出层计算 δ_n 值，
- 5、按背向传播规则的前一层计算 δ_i 值（对每一层都重复计算）
- 6、按下降规则计算权值的变化量（对每个权值都重复计算）



$x_1 = x_2 = 2$ $w_1 = 2, w_2 = -2, w_3 = 4,$ and $w_4 = 0.$