第二章：线性分类器

1. 对二维线性判别函数，$g(x) = x_1 + 2x_2 - 2$

(1) 将判别函数写成 $g(x) = w^T x + w_0$，并画出 $g(x) = 0$ 的图形

(2) 将判别函数写成 $g(x) = a^T y$

参考答案：

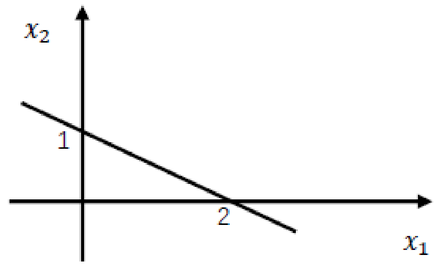
(1) 由$g(x) = x_1 + 2x_2 - 2$ 可知：

$w = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$，$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$，$w_0 = -2$

所以：

$$g(x) = w^T x + w_0 = \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} - 2$$

$g(x) = 0$ 的图形如下：



(2) 由$g(x) = x_1 + 2x_2 - 2$ 可知：

$a = \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix}$，$y = \begin{bmatrix} x_1 \\ x_2 \\ x_0 \end{bmatrix}$，其中$x_0 = 1$，即$y = \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix}$

所以：

$$g(x) = a^T y = \begin{bmatrix} 1 & 2 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix}$$

第三章：决策树学习

3.2　考虑下面的训练样例集合：

实例	分类	a_1	a_2
1	+	T	T
2	+	T	T
3	-	T	F
4	+	F	F
5	-	F	T
6	-	F	T

(a)请计算这个训练样例集合关于目标函数分类的熵。
(b)请计算属性 a_2 相对这些训练样例的信息增益。

参考答案：

(a) 熵的公式：

$$Entropy(S) = -p_+ \log_2 p_+ - p_- \log_2 p_-$$

根据训练样例集合，可知：

$$p_+ = \frac{1}{2}, p_- = \frac{1}{2}$$

代入公式：

$$Entropy(S) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1$$

(b) 信息增益的公式：

$$Gain(S,A) \equiv Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

分析a_2属性可知：

$$a_2 = T, [2+, 2-]$$
$$a_2 = F, [1+, 1-]$$

进一步计算：

$$Entropy(S_T) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1$$

$$Entropy(S_F) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1$$

代入公式：

$$Gain(S,A) \equiv Entropy(S) - \frac{1}{2}Entropy(S_T) - \frac{1}{2}Entropy(S_F)$$
$$\equiv 1 - \frac{1}{2} \times 1 - \frac{1}{2} \times 1$$
$$\equiv 0$$

第四章：

4.3 考虑使用阈值表达式 $w_0 + w_1x_1 + w_2x_2 > 0$ 定义的两个感知器。感知器 A 的权值为：

$$w_0 = 1,\ w_1 = 2,\ w_2 = 1$$

感知器 B 的权值为：

$$w_0 = 0,\ w_1 = 2,\ w_2 = 1$$

请判断以下表达对或错。感知器 A 是 $more\_general\_than$ 感知器 B 的（$more\_general\_than$ 在第 2 章中定义）。

参考答案：

表达正确。

根据书中的公式：$\sum_{i=0}^{n} w_i x_i = w_o + w_1x_1 + \ldots + w_nx_n$

代入权值可得：

A：$w_0 = 1, w_1 = 2, w_2 = 1$

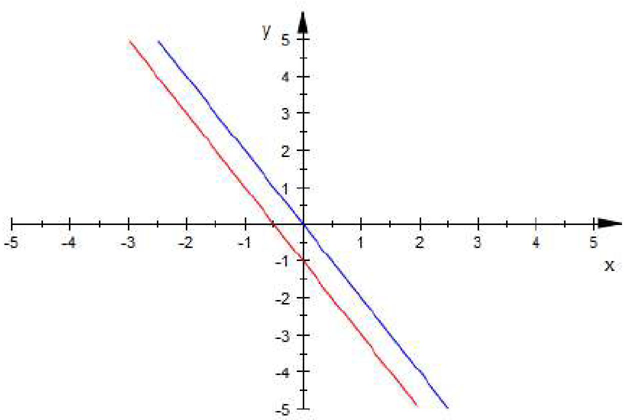
B：$w_0 = 0, w_1 = 2, w_2 = 1$

获得 A，B 的表达式:

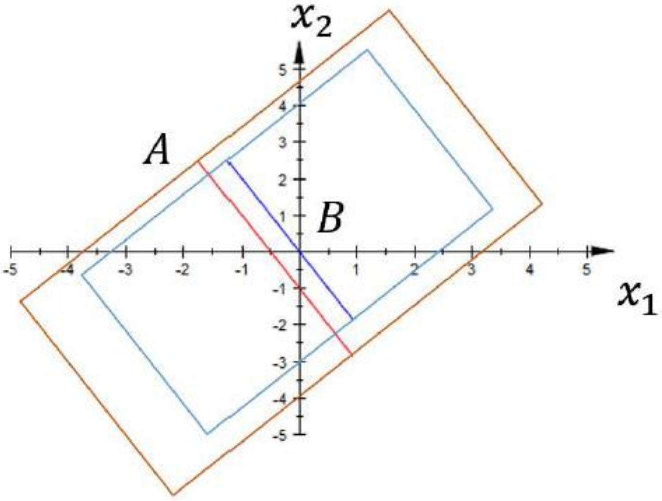
$$A: \sum_{i=0}^{n} w_i x_i = w_o + w_1x_1 + \ldots + w_nx_n = 1 + 2x_1 + x_2$$

$$B: \sum_{i=0}^{n} w_i x_i = w_o + w_1x_1 + \ldots + w_nx_n = 0 + 2x_1 + x_2$$

根据上述线性表达式，可绘制图形，如下所示:



当拥有相同的x_1,x_2时，我们能获得结果：



当A,B有相同的值时，A的范围覆盖B的范围，因此，感知器A more-general-than 感知器B。

4.5 推导输出为 o 的单个单元的梯度下降训练法则,其中:

$$o = w_0 + w_1 x_1 + w_1 x_1^2 + \cdots + w_n x_n + w_n x_n^2$$

参考答案:

$$\begin{aligned} & o = w_0 + w_1 x_1 + w_1 x_1^2 + \cdots + w_n x_n + w_n x_n^2 \\ & \Rightarrow o = w_0 + (w_1 x_1 + \cdots + w_n x_n) + (w_1 x_1^2 + \cdots + w_n x_n^2) \end{aligned}$$

为了推导出单个单元的梯度下降训练规则，我们可以假设训练误差为:

$$E(\vec{w}) \equiv \frac{1}{2} \sum_{d \in D} (t_d - o_d)^2$$

因此:

$$\begin{aligned} \frac{\partial E}{\partial w_i} &= \frac{\partial}{\partial w_i} \left(\frac{1}{2} \sum_{d \in D} (t_d - o_d)^2 \right) \\ &= \frac{1}{2} \sum_{d \in D} \frac{\partial}{\partial w_i} (t_d - o_d)^2 \\ &= \frac{1}{2} \sum_{d \in D} 2(t_d - o_d) \frac{\partial}{\partial w_i} (t_d - o_d) \\ &= \sum_{d \in D} (t_d - o_d) \frac{\partial}{\partial w_i} (t_d - o_d) \end{aligned}$$

对于 $o = w_0 + w_1 x_1 + \cdots + w_n x_n$，结果为: $\sum_{d \in D} (t_d - o_d)(-x_{id})$

对于 $o = w_0 + w_1 x_1^2 + \cdots + w_n x_n^2$，结果为: $\sum_{d \in D} (t_d - o_d)(-x_{id}^2)$

因为:

$$\begin{aligned} & o = w_0 + w_1 x_1 + w_1 x_1^2 + \cdots + w_n x_n + w_n x_n^2 \\ & \Rightarrow o = w_0 + (w_1 x_1 + \cdots + w_n x_n) + (w_1 x_1^2 + \cdots + w_n x_n^2) \end{aligned}$$

所以最终结果为:

$$\sum_{d \in D} (t_d - o_d)(-x_{id} - x_{id}^2)$$

因此，输出为 o 的单个单元的梯度下降训练法则为:

$$w_i = w_i + \Delta w_i$$

其中，$\Delta w_i = -\eta \frac{\partial E}{\partial w_i} = -\eta \sum_{d \in D} (t_d - o_d)(-x_{id} - x_{id}^2)$

第六章：贝叶斯学习

6.1 再次考虑 6.2.1 节中应用贝叶斯规则的例子。假定医生决定对该病人做第二次化验测试，而且化验结果也为正。根据这两次测试，*cancer* 和 $\neg$ *cancer* 的后验概率是多少？假定两个测试是相互独立的。

参考答案：

书中 6.2 的例子，我们能够得到：

$$P(cancer) = 0.008$$

$$P(\neg cancer) = 0.992$$

$$P(+|cancer) = 0.98$$

$$P(-|cancer) = 0.02$$

$$P(+|\neg cancer) = 0.03$$

$$P(-|\neg cancer) = 0.97$$

所以：

$$\begin{aligned}
P(cancer|++) &= \frac{P(++|cancer)P(cancer)}{P(++)} \\
&= P(++|cancer)P(cancer) \\
&= P(+|cancer)P(+|cancer)P(cancer) \\
&= 0.98 \times 0.98 \times 0.008 \\
&= 0.0076832
\end{aligned}$$

$$\begin{aligned}
P(\neg cancer|++) &= \frac{P(++|\neg cancer)P(\neg cancer)}{P(++)} \\
&= P(++|\neg cancer)P(\neg cancer) \\
&= P(+|\neg cancer)P(+|\neg cancer)P(\neg cancer) \\
&= 0.03 \times 0.03 \times 0.992 \\
&= 0.0008928
\end{aligned}$$

归一化结果：

$$\begin{aligned}
P(cancer|++) &= \frac{P(cancer|++)}{P(cancer|++) + P(\neg cancer|++)} \\
&= 0.8959
\end{aligned}$$

$$\begin{aligned}
P(\neg cancer|++) &= \frac{P(\neg cancer|++)}{P(cancer|++) + P(\neg cancer|++)} \\
&= 0.1041
\end{aligned}$$

无监督学习:

聚类——作业

假设挖掘任务是将如下的八个点（用(x, y)代表位置）：

$A_1(2,10)$，$A_2(2,5)$，$A_3(8,4)$，$B_1(5,8)$，$B_2(7,5)$，$B_3(6,4)$，$C_1(1,2)$，$C_2(4,9)$

1、要求聚类为三个簇，假设初始我们选择分别$A_1B_1C_1$为每个簇的中心，用K-均值算法给出：

1）第一轮执行后的三个簇中心

2）最后的三个簇

2、要求聚类为两个簇，采用层次聚类中的合并聚类方法完成聚类。

参考答案：计算过程省略。

1.

1)第一轮执行后的三个簇中心:

（2，10），（6，6），（1.5，3.5）

2) 最后的三个簇:

$\{A_1,B_1,C_2\},\{A_3,B_2,B_3\},\{A_2,C_1\}$

2. 聚类为两个簇:

$\{A_1,B_1,C_2\},\{A_3,B_2,B_3,A_2,C_1\}$