第 1 章：

1.2 挑选一些本书未提到的学习任务。写一段话非正式地加以描述。再尽可能精确地描述出它的任务、性能衡量标准和训练经验。最后，给出要学习的目标函数和它的表示。讨论这个任务设计中考虑的主要折中。

1.3 证明本章描述的 LMS 权更新法则采用了梯度下降方法使误差平方最小化。确切地讲，像文中那样定义误差平方 E。然后计算 E 对权 w_i 的导数，其中假定 $\hat{V}(b)$ 与文中定义的一样，是一个线性函数。梯度下降是通过与 $-\frac{\partial E}{\partial w_i}$ 成比例地更新每个权值实现的。所以，必须证明对于所遇到的每一个训练样例，LMS 训练法则都是按这个比例来改变权值。

第 2 章：

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 $g(x) = a^T y$

第 3 章：

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。

3.3 判断以下命题的正误：如果树 D2 是从树 D1 加工而成的，那么 D1 *more_general_than* D2。假定 D1 和 D2 是表示任意布尔函数的决策树，而且当 ID3 能把 D1 扩展成 D2 时，则 D2 是 D1 加工而成的。如果正确，给出证明；如果错误，举出一个反例（*more_general_than* 在第 2 章中定义）。

第 4 章：

4.2 设计一个两输入的感知器来实现布尔函数 $A \wedge \neg B$。设计一个两层的感知器网络来实现布尔函数 $A\ XOR\ B$。

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）。

4.5 推导输出为 o 的单个单元的梯度下降训练法则，其中：

$$o = w_0 + w_1x_1 + w_1x_1^2 + \cdots + w_nx_n + w_nx_n^2$$

4.7 考虑一个两层的前馈 ANN，它具有两个输入 a 和 b，一个隐藏单元 c，和一个输出单元 d。这个网络有五个权值（$w_{ca}, w_{cb}, w_{c0}, w_{dc}, w_{d0}$），其中 w_{x0} 表示单元 x 的阈值权。先把这些权的值初始化为（0.1, 0.1, 0.1, 0.1, 0.1），然后给出使用反向传播算法训练这个网络的前两次迭代中每一次这些权值的值。假定学习速率 $\eta = 0.3$，冲量 $\alpha = 0.9$，采用增量的权值更新和以下训练样例：

a	b	d
1	0	1
0	1	0

第六章：

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。

6.2 在 6.2.1 节的例子中，为计算癌症的后验概率，通过将 $P(+|cancer)\cdot P(cancer)$ 和 $P(+|\neg cancer)\cdot P(\neg cancer)$ 归一化使它们的和为 1。使用贝叶斯公式和全概率公式（见表 6-1）证明该方法是正确的（即这样的归一化可以得到 $P(cancer|+)$ 的正确值）。

6.3 考虑下面的概率学习算法 *FindG*，它输出一个极大一般化的一致假设（例如，变型空间的某个极大一般成员）。

(a) 给出 $P(h)$ 和 $P(D|h)$ 的分布，以使 *FindG* 保证输出 MAP 假设。

(b) 给出 $P(h)$ 和 $P(D|h)$ 的分布，以使 *FindG* 不能保证输出 MAP 假设。

(c) 给出 $P(h)$ 和 $P(D|h)$ 的分布，以使 *FindG* 保证输出 ML 假设但不是 MAP 假设。

第八章：

8.1 为公式(8.7)中的目标函数的一个距离加权局部线性逼近推导梯度下降法则。

8.2 思考以下为解决局部加权回归中的距离度量的另一种方法。如下建立一个虚拟的训练样例集合 D'：对于原始训练数据集合 D 中的每一个训练样例 $\langle x, f(x)\rangle$，在 D' 中创建出一定数量（可能是分数）的 $\langle x, f(x)\rangle$ 的拷贝，其中拷贝的数量是 $K(d(x_q, x))$。现在训练一个线性逼近来最小化以下误差准则：

$$E_4 = \frac{1}{2}\sum_{x\in D'}(f(x)-\hat{f}(x))^2$$

这里的想法是，对靠近查询实例的训练样例产生较多的拷贝，距离远的拷贝较少。推导出这个误差准则的梯度下降法则。把这个法则表示成在 D 的成员上的求和，而不是在 D' 的成员上求和，并把它与式(8.6)和式(8.7)中的法则进行比较。

第九章：

9.1 为第 3 章中描述的 *PlayTennis* 问题设计一个遗传算法，学习合取的分类规则。精确地描述出其中假设的位串编码和一组交叉算子。

回归分析：

Softmax 推导

无监督学习：