

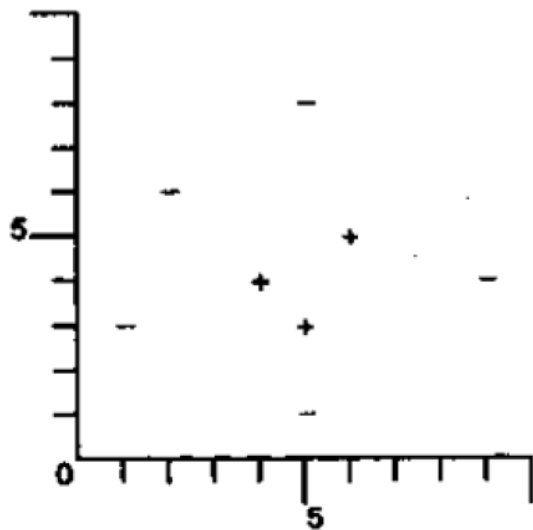
1、假定一实例空间包含 x, y 平面中的整数点，假设集合 H 为矩形集，也就是说，假设的形式为 $a \leq x \leq b, c \leq y \leq d$ ，其中 a, b, c, d 为任意整数。

(a) 考虑对应于下图所示正例(+)和反例(-)集合的变型空间，它的 S 边界是什么? 写出其中的假设并在图中画出。

(b) 变型空间的 G 边界是什么，写出其中的假设并在图中画出。

(c) 假定学习器可提出一个新实例 (x, y) ，并要求施教者进行分类。试给出一个查询，无论施教者怎样分类都能保证减小变型空间。再给出一个不能保证的查询。

(d) 作为施教者，如果想让学习器学习一特定的目标概念(如 $3 \leq x \leq 5, 2 \leq y \leq 9$)，为使候选消除算法完全学习到目标概念，需要提供的训练样本数目最小是多少?



2、如果给定一个无偏的假设空间(即实例的幂集)，学习器将发现每一个未观察的新实例将刚好与变型空间中半数的成员匹配，而不论已使用了怎样的训练样本。证明这一结论。确切地讲，证明对于任意实例空间 X ，任意训练样本集 D 及任意不包含在 D 中的实例 $x \in X$ ，

如果 H 是 x 的幂集，那么在 $VS_{H,D}$ 中有恰好半数的假设将 x 划分为正例，另外半数划分为反例。

3、对于下表的数据

编号	性别	身高(米)	体重(千克)
1	男	1.69	62
2	男	1.75	65
3	男	1.8	75
4	女	1.62	48
5	女	1.68	52
6	女	1.72	63

设计机器学习方法学习通过身高体重推测性别。描述设计的方法，给出要学习的目标函数，并讨论可能存在的问题和改进方法。

4、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。
- (c) 以信息增益为度量标准，画出这些训练样例相应的决策树。