

## 第1章

1.1 编号1为正例, 找到假设空间中与它一致的假设:

10, 12, 14, 16, 58, 60, 62, 64

编号4为反例, 找到假设空间中与它一致的假设

23, 24, 31, 52, 55, 56, 63, 64

编号1除去编号4, 得到 10, 12, 14, 16, 58, 60, 62.

1.3 去噪: 若两个样例所有属性相同, 但取值不同, 则只保留正例,

删除与正例不一致的假设

1.5 ① 信息提取与语义分析

② 提高信息匹配程度

③ 搜索结果排序和个性化推荐

④ 索引构建与优化

⑤ 图片与视频搜索

## 第2章

2.1 应留出的正例反例比应与原数据集相同, 即训练集中 350个正反例  
共有  $C_{50}^{350} \times C_{50}^{150}$  种, 即  $(C_{50}^{350})^2$  种

2.2 10折交叉验证: 每次的测试集都正反例各占一半, 所以错误率为 50%

留一法: 留出正例则训练集中正负比 49:50, 识别为负例,

留出负例则训练集中正负比 50:49, 识别为正例

所以正确率 100%

2.3  $\frac{2 \times P_a \times R_a}{P_a + R_a} > \frac{2 \times P_b \times R_b}{P_b + R_b}$ , 所以当  $P_a = R_a = BEP_a$ ,  $P_b = R_b = BEP_b$  时  
 $BEP_a > BEP_b$ , 假设成立

## 第3章:

3.1 式3.2,  $f(x) = w^T x + b$

情况1:  $y' = wx$  经过原点

情况2: 训练集中每个样本都减去某一样本。

3.6 用适当的映射方法使其在更高维度上可分,如KLDA和线性判别分析

3.9 对OVR, MvM来说,由于对每个类进行了相同的处理,其拆解出的二分类任务中类别不平衡的影响会相互抵消,因此通常不需专门处理