HW1

1.2 析合范式

合取若干个析取范式：为真的情况取和 总共的析取范式有 $3*4*4+1=49$ 个，加一为没有为真的情况，表示为空。 所有的特征组合有 $2*3*3=18$ 种，所以最大的特征假设数为$2^{18}=262144$ 种。

这里简单给出计算代码：48种析取式子用18位的整型数字表示，1表示对应的特征组合为True，0表示为False。从48个析取范式种挑选k个C_{48}^{k}种组合，加入一个set集合中，最后打印set集合的大小.

states:

```
#这里的sto为存储48种18位的析取范式
def findK(sto, i, j, num):
    res = set()
    #选择长度为零的时候返回0,
    if num == 0:
        res.add(0)
        return res

    for k in range(i, j - num):
        #第num个数选定为 sto[k],在剩下的k+1到j里面选取num-1个
        reset = findK(sto, k + 1, j, num - 1)
        for var in reset:
            res.add(var | sto[k])
    return res

for k in range (1, 48):
    s = set()
    res = findK(sto, 0, N, k)
    print(k, len(res))
```

```
k 种数(析取范式不含为空时)
1 48
2 879
3 8223
4 40911
5 112962
6 193998
7 233640
......
```

```
k 种数(析取范式含为空时)
1 49
2 897
3 8367
4 41580
5 114990
6 198444
7 241548
......
```

另，网上流传的一份答案。作者使用栈的方式打算遍历 2^{48} 种所有组合，来分析各种k组合的情形。但是栈只能记录一个数值，无法保证所有组合的去重。这也就是计算结果出现组合数大于假设空间 2^{18} 的原因。

1.3 若数据包含噪声，则假设空间中有可能不存在与所有训练样本都一致的假设。在此情形下，试设计一种归纳偏好用于假设选择。

1. 对于样本特征相同，标签不同的数据，可以认为标签类别为较多的一类为真
2. 如此类样本数目不多，可以直接剔除此类数据

言之有理即可