Lecture1作业

1，现在很多火车站都可以持身份证和车票“刷脸”进站，张同学拿着自己的身份证进站时机器却识别错误，从模式识别的角度看机器难以识别他的原因是：

（a）不存在模式； （b）没有训练样本；

（c）不同类的模式相似度高； （d）同一类的模式差异性大。

2，某研究人员用带有标签的数据库训练其目标识别算法，希望用在自动驾驶中对路况进行场景理解和行人检测，请问这个算法需要去解决哪种类型的学习问题：

（a）有监督学习； （b）无监督学习；

（c）半监督学习； （d）强化学习。

3，如果一个好的学习算法通过训练样本集在解空间（假设空间）集合中找到一个最优解，使得其对所有训练样本都能够实现正确的模式识别，以下哪种说法正确：

（a）目标函数是已知的；

（b）所有训练样本都是线性可分的；

（c）训练样本集上全部分类正确，不能代表测试时也能正确；

（d）好的学习算法能够容许训练样本存在错误的标签。

4，某个电影网站每名网络用户都可以给电影打分，用户有id号，电影也有id号，网站已积累了10万个用户对所看电影喜爱程度的打分值，并将电影的特征如喜剧片、动作片、爱情片、…、明星甲、明星乙等与观众对这些特征的喜好匹配组合到一起设计了推荐系统，给每部电影一个百分制的评分结果。请用上述这些元素去构造一个学习问题。

5，有一数据集共有2000张花卉图片，其中1400张是玫瑰，300张是月季，300张是蔷薇。某同学设计了一个玫瑰识别算法，从2000张图片中识别出1000张图片为玫瑰，但实际上其中只有600张是玫瑰，另外300张是月季、100张是蔷薇。请计算分类正确率 (Accuracy)、分类错误率 (Error rate)、分类精度 (Precision)、召回率 (Recall)、F1分数 (F1 Score)。如果该同学的算法把2000张图片都识别成玫瑰，请再次计算出上述指标。