1. 算法原理

1.1最小距离

最小距离分类，是指求出未知类别[向量](https://baike.baidu.com/item/%E5%90%91%E9%87%8F/1396519)到要识别各类别[代表](https://baike.baidu.com/item/%E4%BB%A3%E8%A1%A8/3357884)向量中心点的距离，将未知类别向量归属于距离最小一类的一种图像分类方法。

最小距离分类法是[分类器](https://baike.baidu.com/item/%E5%88%86%E7%B1%BB%E5%99%A8/3317404)里面最基本的一种分类方法，它是通过求出未知类别[向量](https://baike.baidu.com/item/%E5%90%91%E9%87%8F/1396519)X到事先已知的各类别（如A，B，C等等）中心向量的距离D，然后将待分类的向量X归结为这些距离中最小的那一类的分类方法。

* 1. KNN

KNN算法的核心思想是如果一个样本在特征空间中的k个最相邻的样本中的大多数属于某一个类别，则该样本也属于这个类别，并具有这个类别上样本的特性。该方法在确定分类决策上只依据最邻近的一个或者几个样本的类别来决定待分样本所属的类别。 kNN方法在类别决策时，只与极少量的相邻样本有关。

算法流程：

1. 准备数据，对数据进行[预处理](https://baike.baidu.com/item/%E9%A2%84%E5%A4%84%E7%90%86)

2. 选用合适的数据结构存储训练数据和测试元组

3. 设定参数，如k

4.维护一个大小为k的的按距离由大到小的[优先级队列](https://baike.baidu.com/item/%E4%BC%98%E5%85%88%E7%BA%A7%E9%98%9F%E5%88%97)，用于存储最近邻训练元组。随机从训练元组中选取k个元组作为初始的最近邻元组，分别计算测试元组到这k个元组的距离，将训练元组标号和距离存入优先级队列

5. 遍历训练元组集，计算当前训练元组与测试元组的距离，将所得距离L 与优先级队列中的最大距离Lmax

6. 进行比较。若L>=Lmax，则舍弃该元组，遍历下一个元组。若L < Lmax，删除优先级队列中最大距离的元组，将当前训练元组存入优先级队列。

7. 遍历完毕，计算优先级队列中k 个元组的多数类，并将其作为测试元组的类别。

8. 测试元组集测试完毕后计算误差率，继续设定不同的k值重新进行训练，最后取误差率最小的k 值。

* 1. SVM

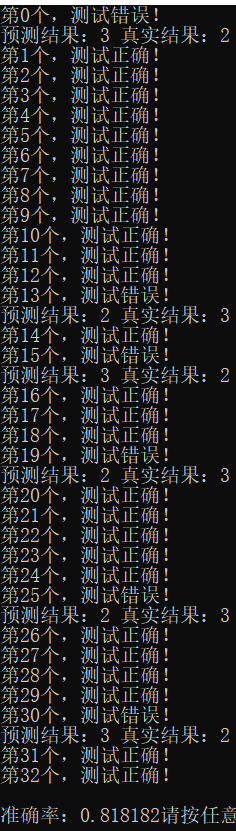
支持向量机（Support Vector Machine, SVM）是一类按[监督学习](https://baike.baidu.com/item/%E7%9B%91%E7%9D%A3%E5%AD%A6%E4%B9%A0/9820109)方式对数据进行[二元分类](https://baike.baidu.com/item/%E4%BA%8C%E5%85%83%E5%88%86%E7%B1%BB/15635322)的广义线性分类器，其[决策边界](https://baike.baidu.com/item/%E5%86%B3%E7%AD%96%E8%BE%B9%E7%95%8C/22778546)是对学习样本求解的最大边距超平面。

SVM使用铰链损失函数计算经验风险并在求解系统中加入了正则化项以优化结构风险（，是一个具有稀疏性和稳健性的分类器。SVM可以通过[核方法](https://baike.baidu.com/item/%E6%A0%B8%E6%96%B9%E6%B3%95/1683712)进行非线性分类，是常见的核学习方法之一。

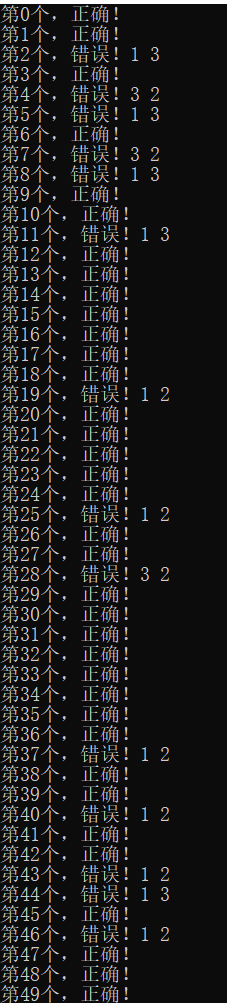
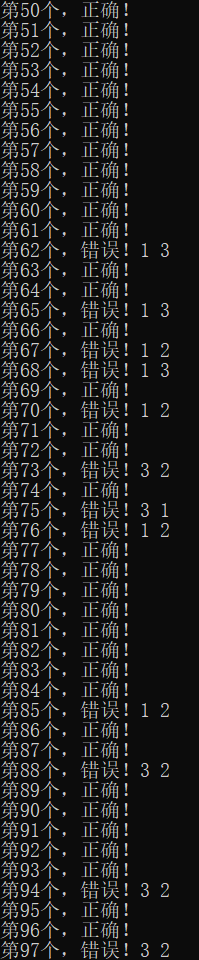
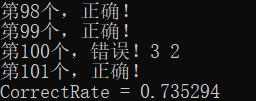
常见的核函数有：线性核，多项式核，径向基函数和，拉普拉斯核，Sigmoid核等。

1. 分类过程

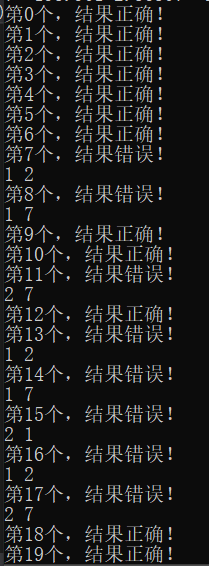
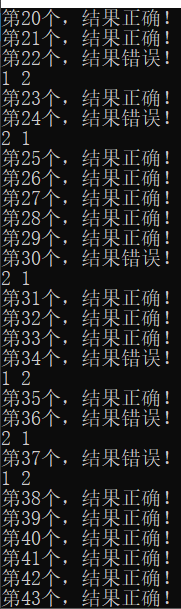
2.1最小距离



2.2KNN

2.3 SVM

1. 结果图形化比较

最小距离的准确率最高，KNN次之，SVM最低。三种方法各有其适用的场合，相对而言，最小距离最简单，适用范围最小。