# 报告

## 一：算法原理

在本次实验共用到了3中方法，分别是：最小距离、K近邻、SVM。

1. 最小距离

最小距离分类法是分类器里面最基本的一种分类方法，它是通过求出未知类别向量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。在计算距离是时有多种方式，如，欧式距离、马氏距离、曼哈顿距离。本次实验使用的是欧式距离来计算两向量之间的距离。

1. K近邻

过度量“待分类数据”即测数据数据和“类别已知的样本集”即训练集的距离对样本进行分类。它根据离得最近的点中找最近的k个点以这几个点的类别进行投票以最后比例确定新的点类别的更大可能性。本质：如果两个样本足够的相似的话他们就有更高的可能性属于同一类别。而这个相似度是通过距离进行衡量的，本次使用的是欧式距离。

1. SVM

SVM即可处理线性可分问题，也可以通过引入核函数处理非线性分类问题。SVM 的核心思想是尽最大的努力使分开的两个类别有最大间隔，这也是有别于逻辑回归的地方，这样才使得分割具有更高的鲁棒性。而且对于未知的新样本才有很好的分类预测能力。所以SVM让离分隔面最近的数据点具有最大的距离。所以SVM的优化目标就变为下式的形式：

目标函数是二次的，约束条件是线性的，所以它是一个凸二次规划问题，现在对于凸二次规划问题有着很成熟的解决方法。为了核函数的引入，该问题是通过使用拉格朗日对偶问题来解决这个问题。

## 二：程序流程

1. 最小距离

程序的流程如下图，在测试样本的类别判断上，当样本到哪个类别中心的距离最小，则该测试样本就被归为该类别当中。



1. K近邻



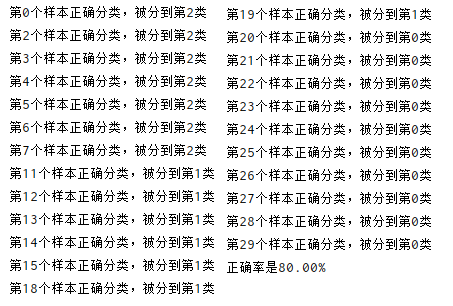
在K近邻方法中，当计算完一个测试样本到所有训练样本之间的距离后，通过选择前k最近距离，通过统计此时所选出的训练样本中各类别出现的次数，则测试样本的类别为出现次数最多的类别。

1. SVM

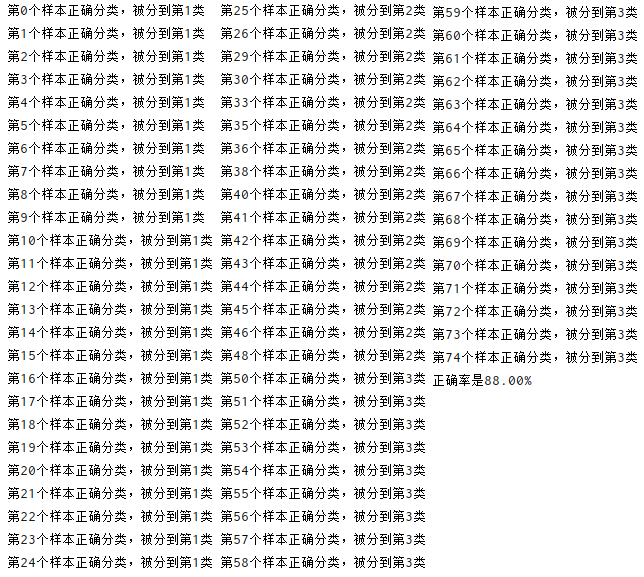


## 三：结果对比

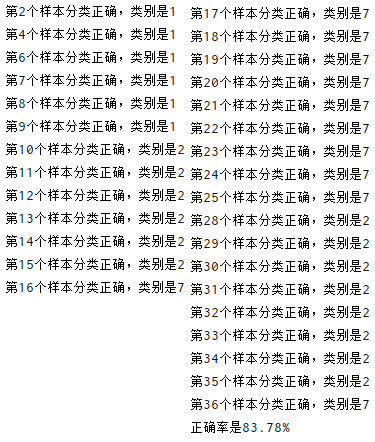
1. 最小距离的分类结果如下：



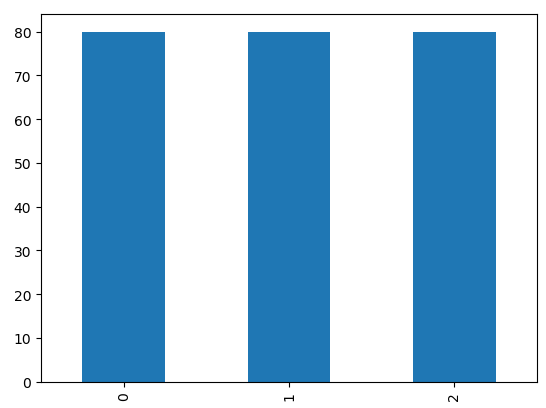
1. k=6近邻的分类结果如下:



1. SVM的分类结果如下:



1. 准确率对比



0：最小距离准确率为：80.00%

1：k近邻准确率为：88.00%

2：SVM准确率为：83.78%