基于支持向量机的图像分类算法

赵思源（919101960141 机械工程学院）

**摘要**：为了实现对图像数据的自动分类，设计了图像分类算法，对图像的特征与分类理论展开分析与研究。首先提取图像中的颜色与轮廓信息，接着使用主成分分析进行降维处理，再对提取信息进行可分离性分析选择软硬间隔、核函数种类并使用网格搜索法调节参数，最终实现分类，并与神经网络与最近邻分类算法比较效果。实验表明此方法基于cifar10数据集的十分类问题正确率为56.8%。相较于随机分类正确率有极大提高，具有运行速度快，准确率高的特点。

**关键字**：图像分类、主成分分析、支持向量机、神经网络、最近邻算法

1.引言

伴随着计算机与模式识别技术的高速发展，图像分类在愈来愈多的领域都有着广泛的应用，如：安防领域的人脸识别和智能视频分析，交通领域的交通场景识别，互联网领域基于内容的图像检索和相册自动归类，医学领域的图像识别等。．可以说，图像分类技术已经应用于人们日常生活的方方面面，计算机自动分类与检测技术也在一定程度了减轻了人的负担，改变了人类的生活方式。

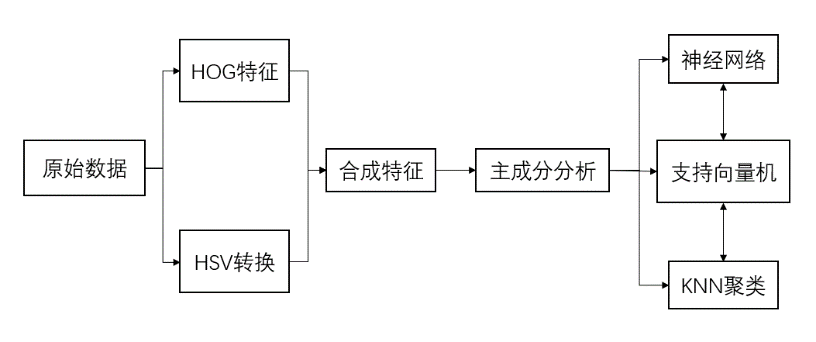
图像分类作为计算机视觉重要的基础问题之一，是物体检测、图像分割、物体跟踪、行为分析、人脸识别等其他高层视觉任务的基础：欲对实际复杂场景进行自动分析与理解，首先就需要确定图像中存在什么物体，因此研究准确且高效的图像分类算法有着重要的理论意义和实际意义。

2.研究现状

目前应用较为广泛的图像分类技术主要分为传统的诸如K-最近邻算法、支持向量机机器学习方法和目前高速发展的卷积神经网络模型，自LeNet5[1]网络提出以来，在 AlexNet[2]引起了大量关注后，应用于图像分类的卷积神经网络技术高速发展，逐渐诞生了GoogLeNet[3]模拟神经网络对猫的关键特征进行自行学习并分类、ResNet[4]解决了增加深度带来的退化问题、SENet[5]方便集成并提升性能等一系列优秀的模型。

但是，由于神经网络有大量的参数，容易在训练数据集规模较小时发生在训练集上准确率很高，在测试集上效果不尽人意的过拟合问题。且计算资源有限，训练网络需要花费很长的时间。支持向量机的优势由此体现：作为1992年和1995年发表的技术[6]，支持向量机在当时并没有立刻引起很大的效应，但其与传统模式识别以及人工神经网络相比，能够在样本数相对较少，特征维数高的情况下仍然取得很好的推广能力。结合模糊支持向量机、基于决策树的支持向量机、分级聚类的支持向量机等理论改进[7]，支持向量机在图像领域很常见，因为该领域很容易体现支持向量机的优势。

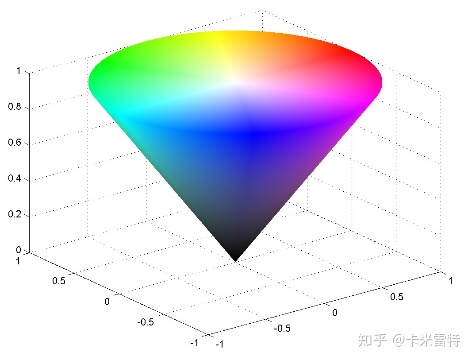
3.图像分类方法

本图像分类算法的流程如下：

3.1图像特征提取

在计算机中，图像以有序的多维数组形式储存，即使一张很小的图片，也需要大量的存储空间保存图像的全部信息。直接以图像的存储矩阵数据进行各类计算会造成大量的资源浪费与计算困难。因此，对图像特征进行提取可以用较低的维数概括图像的信息。使用图像特征而非原始图像数据为数据的存储和使用带来了极大的方便。

3.1.1色彩空间(HSV)转换

 HSV是一种比较直观的颜色模型，不同于传统的RGB颜色通道，HSV颜色的参数为色调(H)，饱和度(s)，明度(v)，此形式下的图像更接近人们对彩色的感知经验，可以非常直观的表达颜色的色调、鲜艳程度与明亮程度；同时，相比RGB空间，在此空间下的图像颜色特征更加突出，方便根据图像的颜色特征进行识别、比对与分类。

图像：HSV空间

为方便后续图像处理，用R、G、B代指原始图片像素点的各通道强度，并定义：



使用以下公式对原始数据集进行从RGB空间到HSV空间的转换：



3.1.2梯度直方图(HOG)特征

HOG特征是一种在计算机视觉和图像处理中用来描述图像梯度的特征，能够很好地描述图像中边缘部分的方向密度。

首先对原始的彩色图像进行灰度化处理并对颜色空间进行标准化处理，由此调节图像的对比度，起降低图像光照变化影响、消除噪声的作用。然后使用sobel算子依次计算划分后子图像在水平和竖直两方向的梯度，并由此计算图像梯度的大小与方向：



由于局部光照的变化以及前后景对比度的变化使得梯度强度的范围很大，因此接下来对图像的梯度进行归一化处理，通过将图像分为若干单元格并统计单元格中梯度之和，再将各个单元格组成大的、空间上连通的区间，将该区间中所有单元格的梯度串联起来，由此获得图像的HOG特征。

图像：在原始图像上用白色箭头标出HOG特征，（基于cifar10数据集第41和第141张照片）

3.1.3合成特征向量

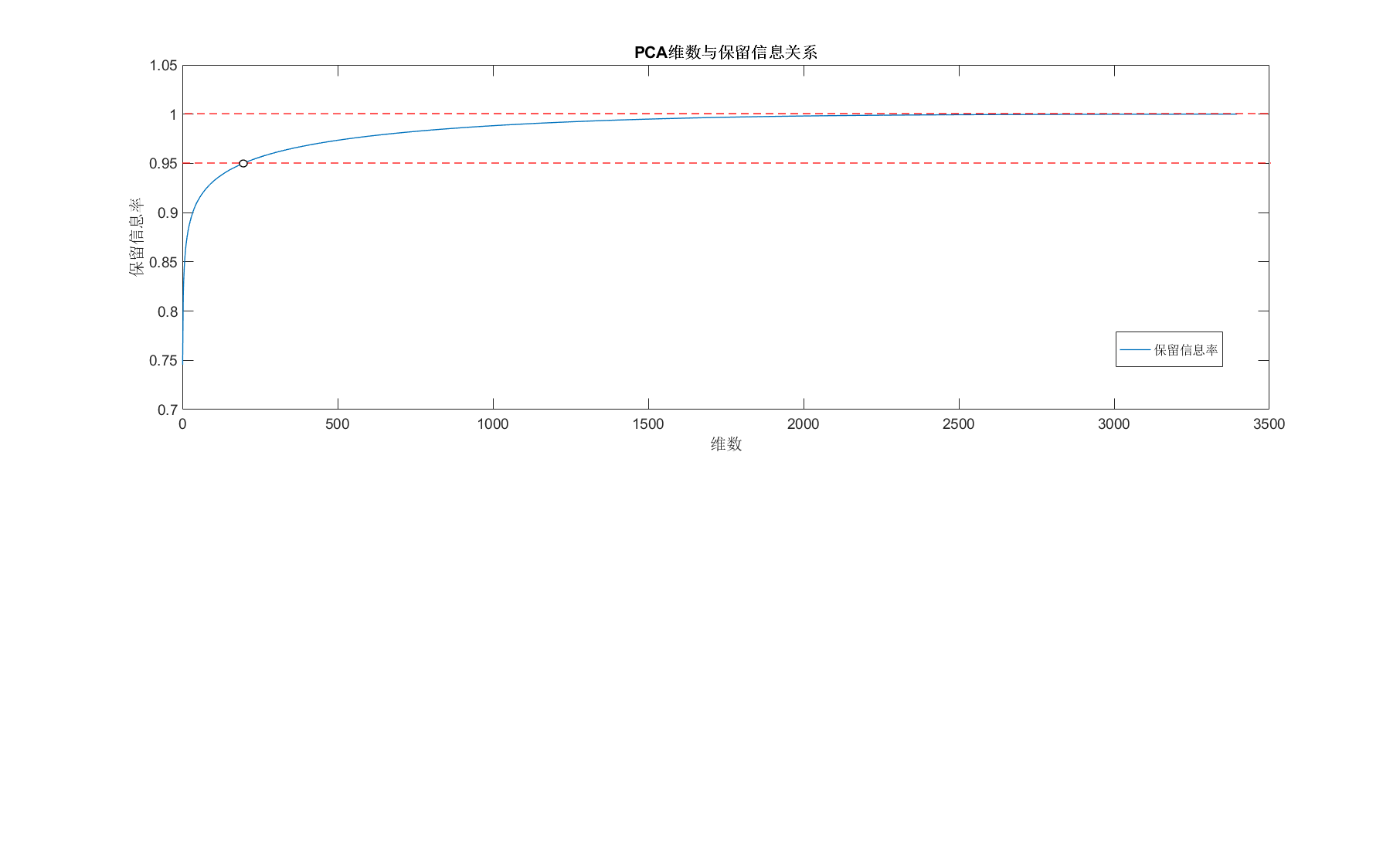
在通过上述流程计算获得图像的HSV特征用以代表原始图像的颜色特征和HOG特征代表原始图像的边缘信息后，将每张图片的HOG和HSV特征向量连接起来并归一化，构成一个总的数据集的向量矩阵，该矩阵的每一行是单个图像的HSV+HOG特征。使用此向量矩阵代替原始的图像RGB特征，根据以往实验经验，HSV+HOG特征相比RGB，更好的反映了图像本身的特征信息，能够有效提高后续图像分类的正确率。

3.2 主成分分析法(PCA)降维

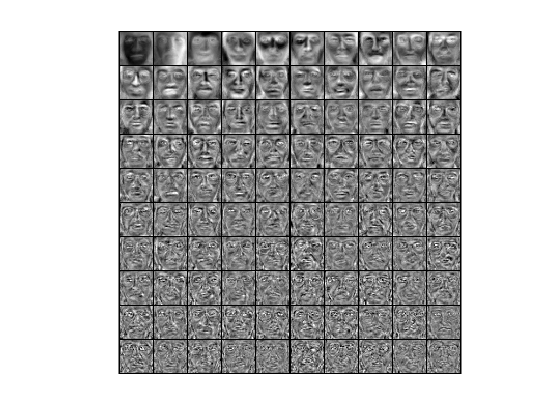
主成分分析法可以在保留原始数据的大量特征信息的前提下降低图像的维度，方便后续处理。假定原始矩阵为M，根据以下公式：



将原始矩阵进行奇异值分解，中间的非负实数对角矩阵对角线上的矩阵即为原始矩阵M的奇异值，通过奇异值的大小可以确定各维度包含原始数据信息量的大小。



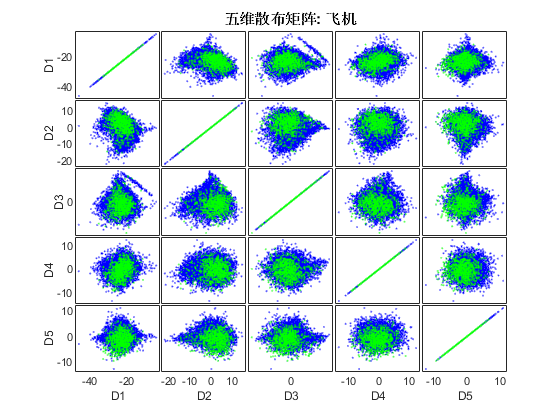
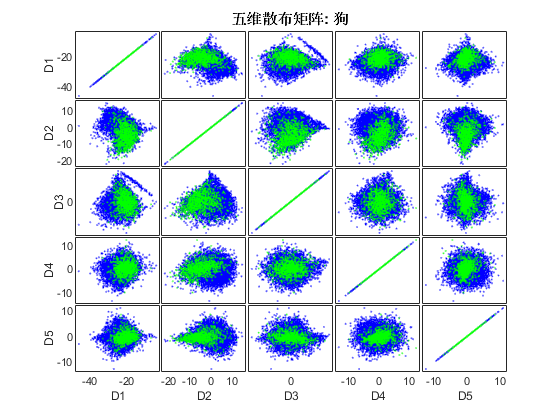
图像：pca维数与包含信息量关系（基于cifar10数据集）

图像：人脸各维度包含信息，左边为原始图像，右边为特征脸，由观察可知单个特征脸所包含的信息随着对应维数的增加越来越少（基于YaleB数据集前一百张人脸）

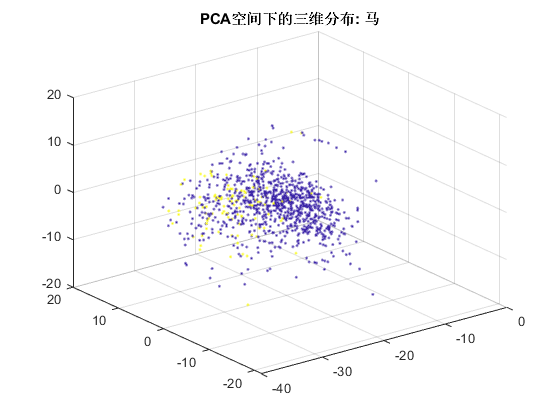
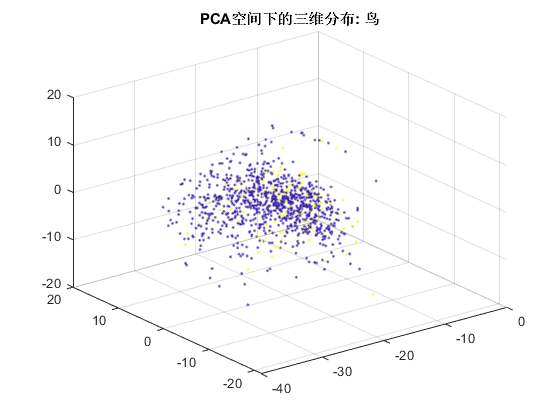
维数越多，信息利用率也越高，但处理起来也更加困难，因此需要选择合适的维数。因此选择包含95%有效信息的维数。

完成PCA主成分分析法降维维数确定后，利用奇异值分解获得的酋矩阵各列代表的主成分方向，可将原始数据进行坐标变换，投影至PCA空间，完成主成分分析，将原始数据中的绝大部分有效信息成功提取。



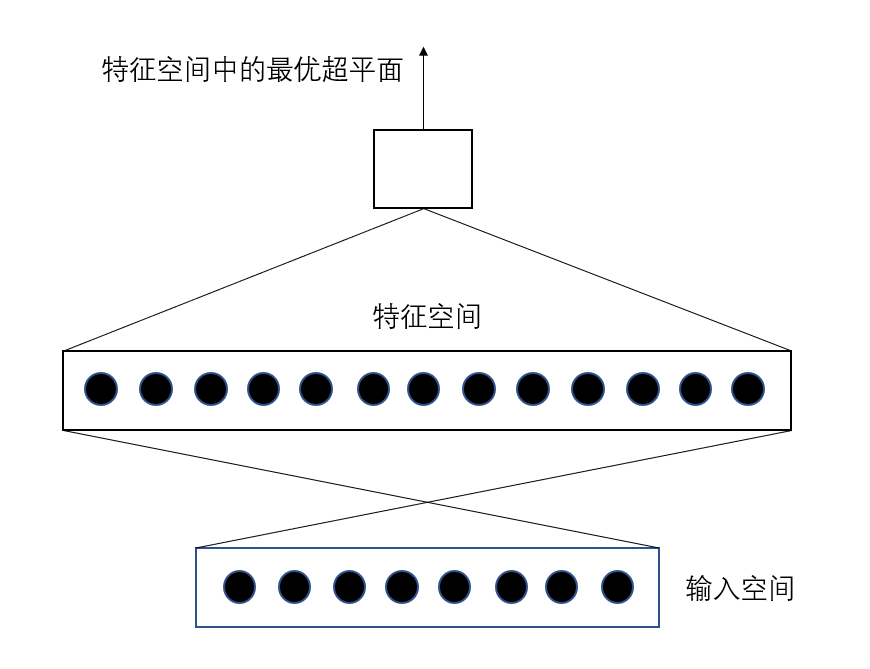
图像：PCA分析后，同类数据的五维散布矩阵（基于cifar10数据集）

3.3 支持向量机构建

对当前图像数据可视化可看出，由于各类图像数据混杂在一起，使用硬间隔支持向量机将会导致计算困难，因此采用软间隔支持向量机进行分类。

图像：三维空间下马与鸟类图像位置分布，黄点为该种类图片，紫点为其他图片（基于cifar10数据集）

3.3.1核函数选择

在支持向量机中，为了保证原始数据可分离，有时需要使用核函数方便的将当前的数据映射至高维空间中去，直接映射会导致其计算的复杂度大幅增加，因此需要使用核函数对原始数据进行处理。

图片：非线性变换实现非线性向量机

设当前输入数据的正负样本分别为，核函数为,假定函数Φ能够将样本特征映射至高维空间，则核函数应满足：



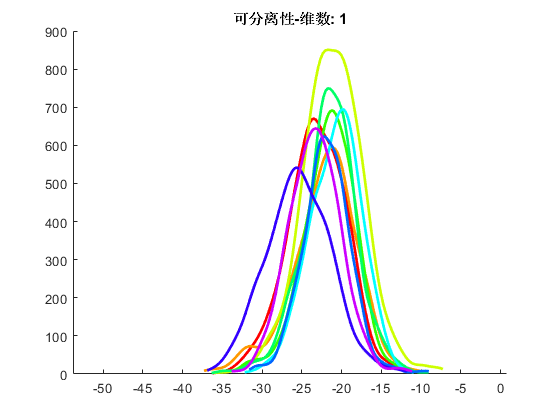
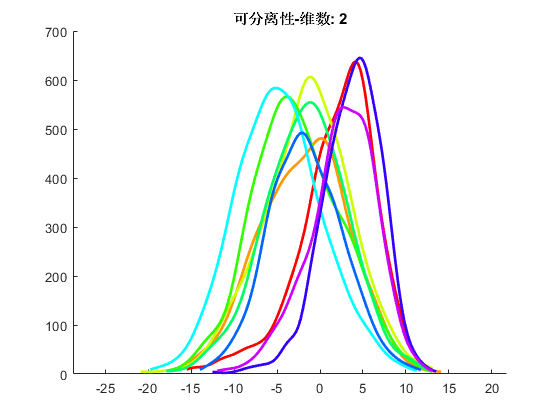
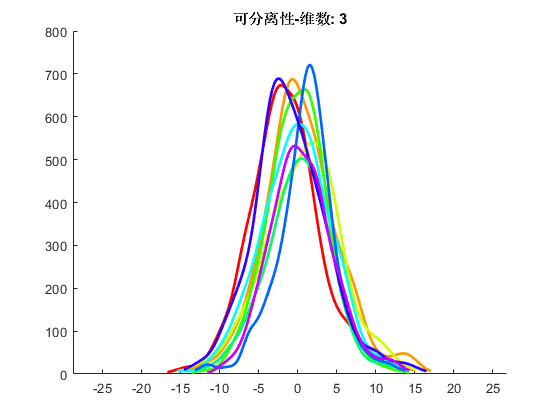
目前主要使用的支持向量机核主要有以下三种：

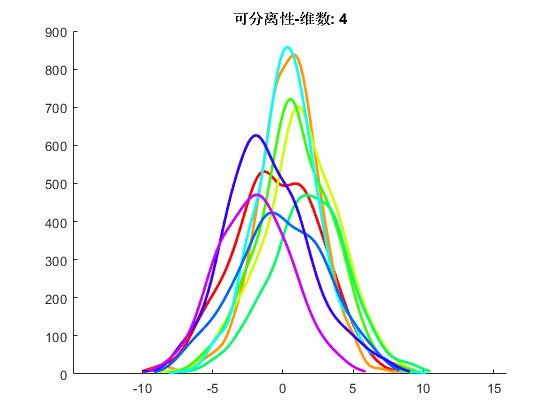
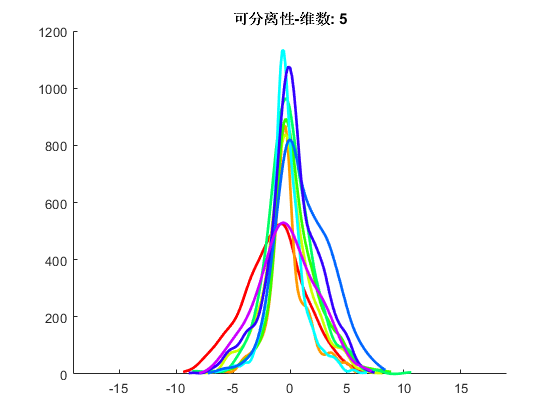
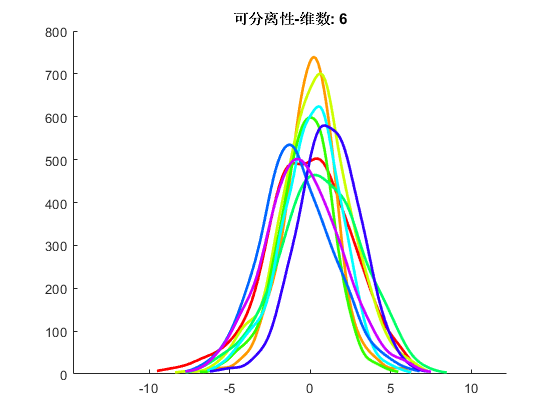
（1）线性核。

线性核形如



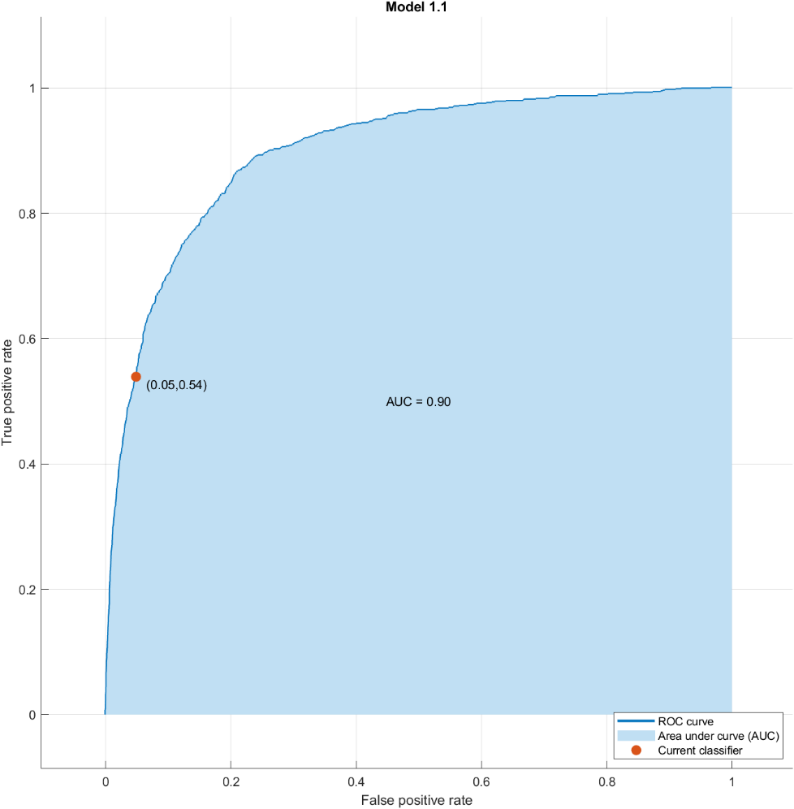
线性核相当于并未对输入空间进行向高维的变换，因其形式简单，速度较快，故首先尝试使用线性核分类。

首先做出支持向量机输入空间数据的统计分布图，用以判断其是否具有一定的可分离性。



图片：输入空间前六维数据的统计分布

观察得知，输入空间中数据的特征差异较大，有一定的线性可分性，故尝试使用线性核。经测试，线性核基本达到了分类的要求，但仍有提升空间。



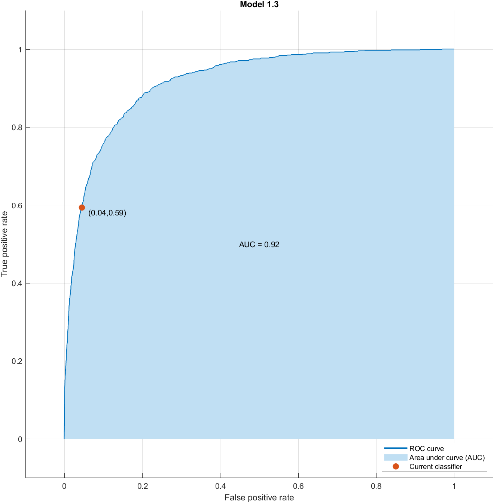
图片：线性核分类的混淆矩阵，总正确率为49.8%

（2）多项式核。

多项式核形如：



能够将低维数据映射至高维，但其计算较慢，且参数复杂，难以调试。尝试使用此核支持向量机对数据进行分类，准确率较线性核有所上升，但其时间成本过高，故不采用此核。



图片：三次多项式核的分类效果，总正确率52.3%

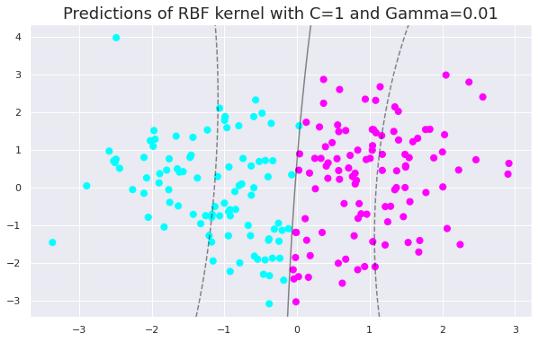
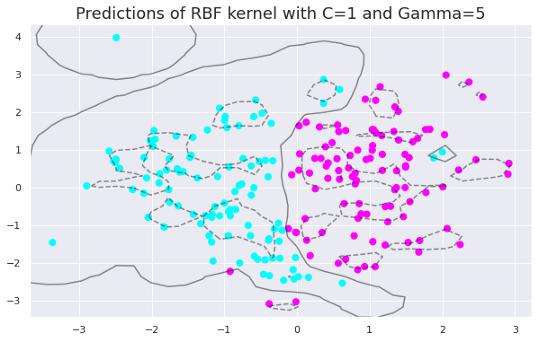
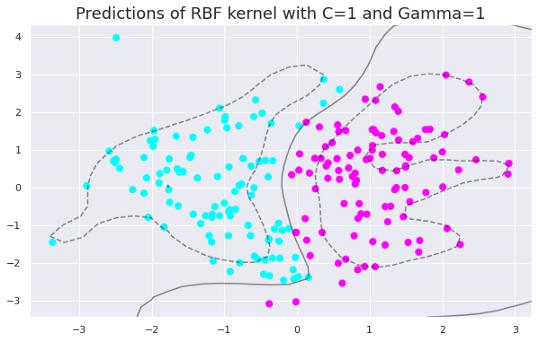
（3）高斯核。

高斯核形如：



高斯核函数通过高斯公式，同样能够实现将数据映射至高维，且计算量较小，无论样本大小都能获得较令人满意的性能，初步测试其分类性能优于线性分类，故选用高斯核作为此图像分类任务的核。

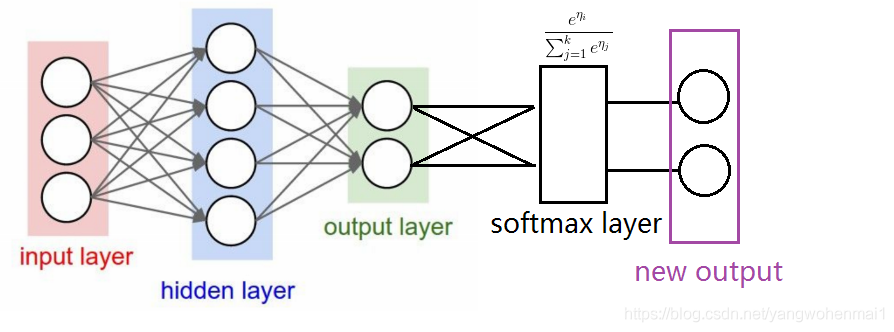
3.3.2误差惩罚参数选择

将高斯核函数的惩罚参数记为c，核函数中的参数记为γ，c和γ的取值均会对分类的效果产生巨大的影响，故采用网格搜索法【8】对最佳c和γ进行遍历搜索，选取最佳组合作为支持向量机的参数。

图片：不同的γ导致的不同聚类结果

3.4 神经网络构建

神经网络由大量的人工神经元联结进行计算。大多数情况下人工神经网络能在外界信息的基础上改变内部结构，是一种自适应系统。现代神经网络是一种非线性统计性数据建模工具，常用来对输入和输出间复杂的关系进行建模，或用来探索数据的模式。

为方便比对支持向量机分类效果，本实验同样构建了简单的神经网络，网络的结构如下：

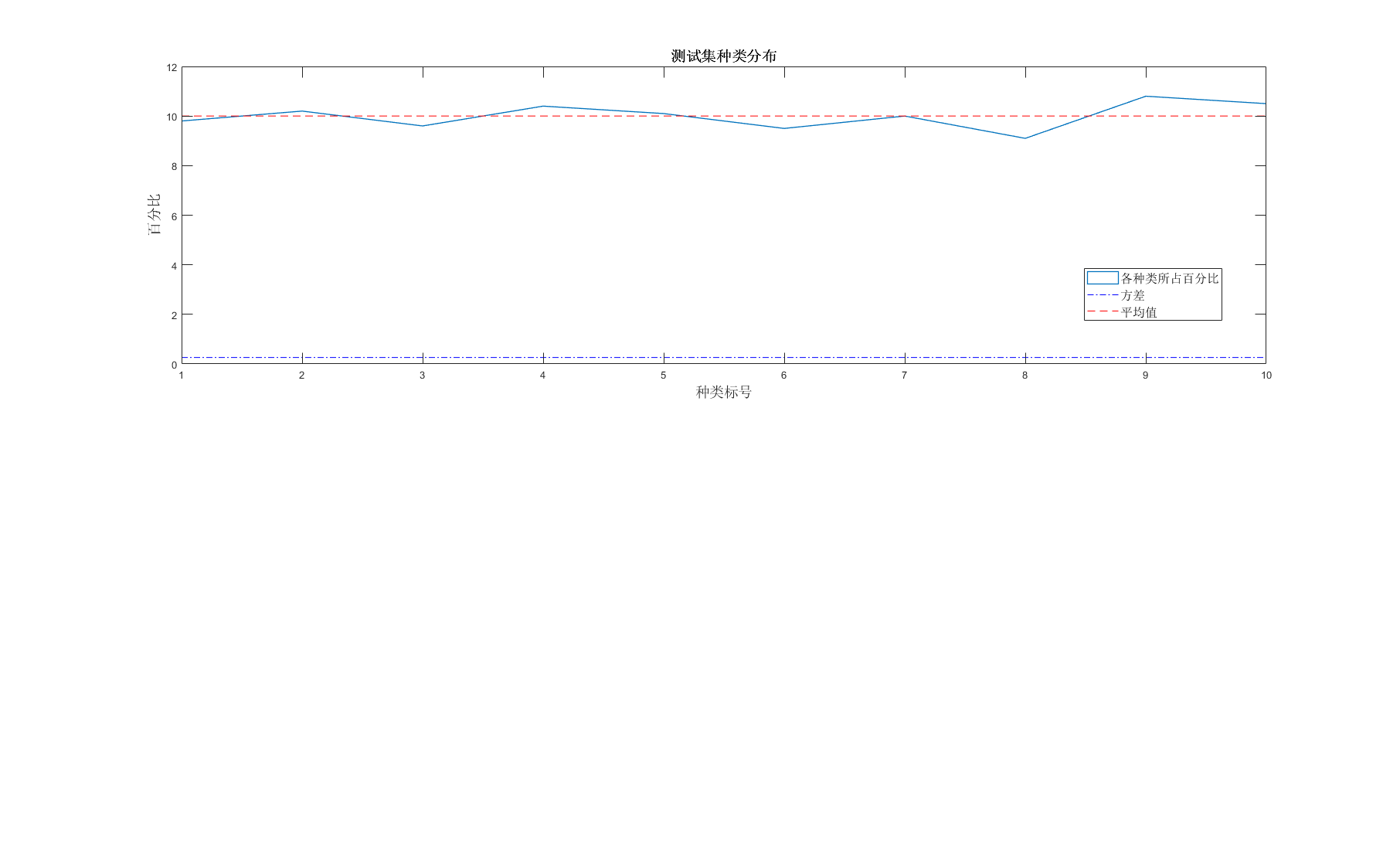
其hidden layer仅包含一层全连接层。经测试，该网络结构能够实现对图像的分类

4.实验结果

4.1实验数据

以cifar10数据集的batch1进行测试，此数据集共10000张32\*32大小的彩色图片，这些图片可被分为10类。

4.2实验设置

将90%的图像用作训练，10%图像用作测试。使用KNN聚类算法、神经网络和支持向量机分别对其进行训练，并使用测试集测试，比较三者准确率差异

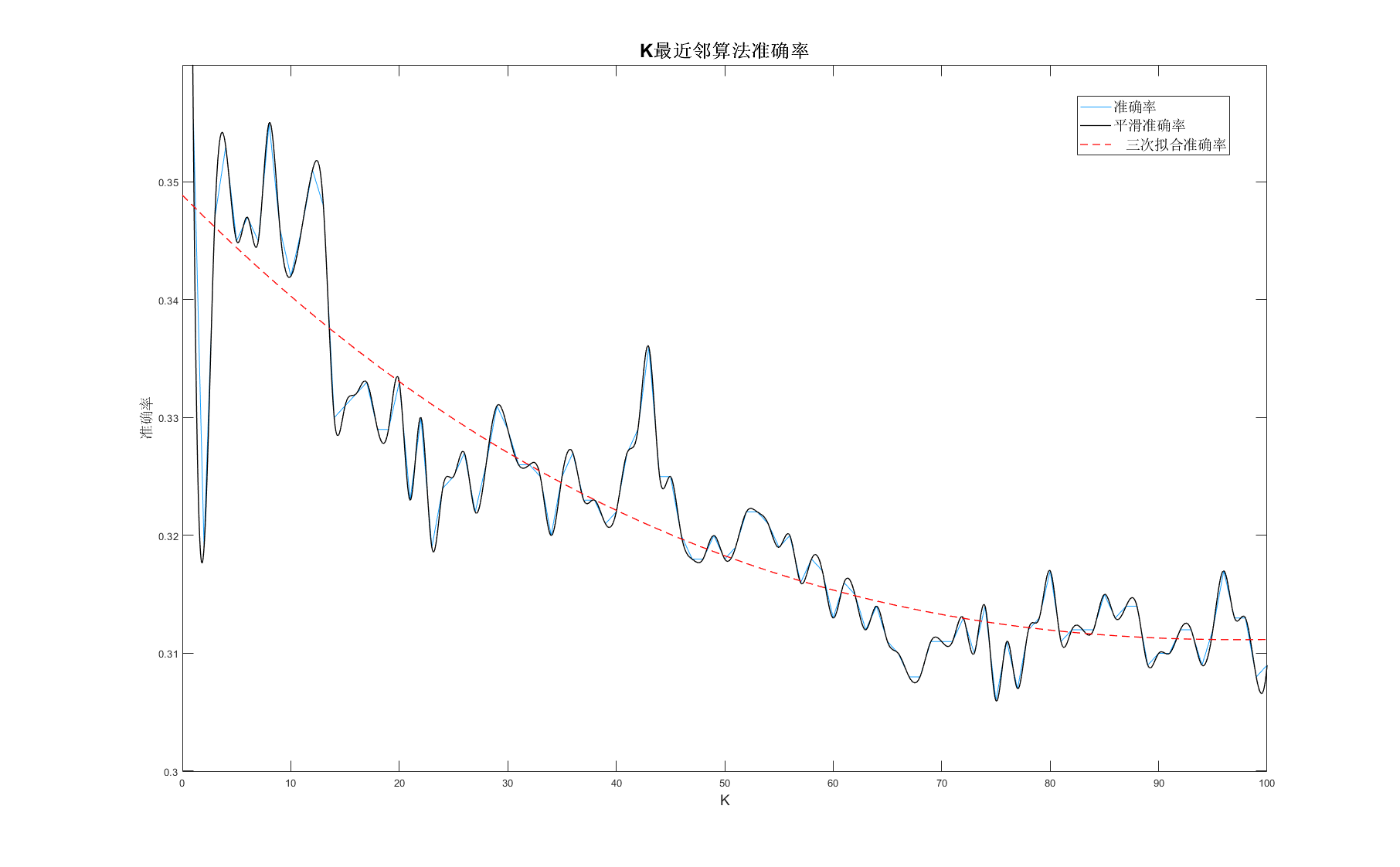
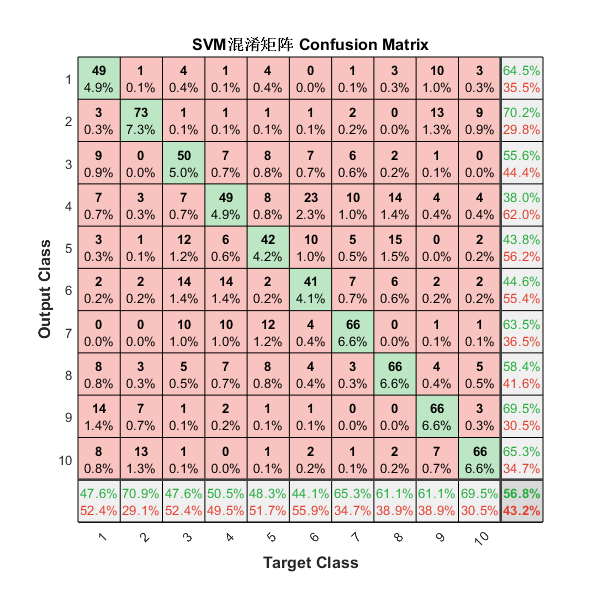
图片：测试集数据种类分布

4.3实验结果：

使用高斯核支持向量机获得混淆矩阵如下：

图片：分类混淆矩阵

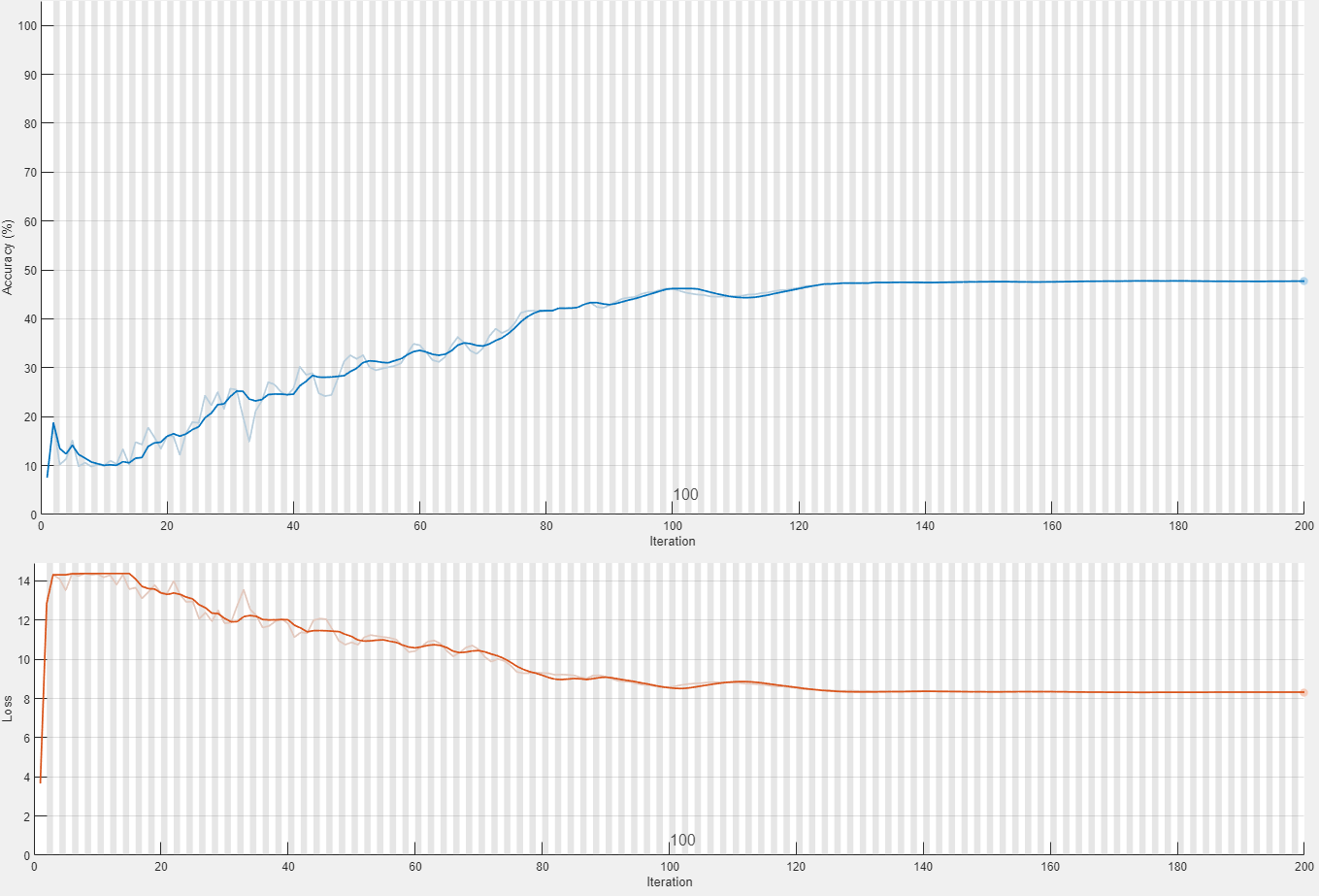
使用测试集测试，分类正确率为56.8%。

 4.3.1与k-最近邻分类对比

图片：KNN分类准确率（基于cifar10数据集）

KNN分类方法的准确率与K值的选择相关，在一定范围内，识别的准确率随着K值的增大而减小。超过某阈值后，KNN分类的准确率会逐渐提升，但此时由于考虑点过多，KNN分类已几乎失去其作用。

选用合适的k值进行分类，得到的正确率位于30%-35%左右，与支持向量机相比，即使是最好的情况，正确率也明显低于分类最简单直接的线性核支持向量机分类效果。

4.3.2与神经网络对比

图：神经网络的变化（上图为准确率，下图为loss）

最终训练获得的神经网络正确率为44.6%，虽可通过引入残差网络等多种方式改善正确率，但计算成本过高。

5.结束语

基于支持向量机的图像分类算法通过多个支持向量的核函数内积加权和实现了各种非线性分布面，完成了图像的分类任务。对于此任务，该分类算法相较于传统的KNN分类算法，它的正确率有了大幅提升；相较于高准确率的神经网络，它的计算成本较低。因此，对于现实中的各种问题，模型与参数的选择都需要与样本和问题的复杂性相适应。支持向量机作为一种历史悠久的分类算法，充分考虑应用领域的相关知识后仍可以实现较好的分类效果，在信息量日渐庞大的当下，能够在语音识别、文字识别、图像识别等领域充分发挥其作用。

参考文献：

* [1]Y. LeCun,L. Bottou,Y. Bengio,P. Haffner. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE,1998,86(11).
* [2]Christian Szegedy,Wei Liu 0015,Yangqing Jia,Pierre Sermanet,Scott E. Reed,Dragomir Anguelov,Dumitru Erhan,Vincent Vanhoucke,Andrew Rabinovich. Going Deeper with Convolutions.[J]. CoRR,2014,abs/1409.4842.
* [3]Kaiming He,Xiangyu Zhang,Shaoqing Ren,Jian Sun 0001. Deep Residual Learning for Image Recognition.[J]. CoRR,2015,abs/1512.03385.
* [4]Hu Jie,Shen Li,Albanie Samuel,Sun Gang,Wu Enhua. Squeeze-and-Excitation Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2019,42(8).
* [5]Corinna Cortes,Vladimir Vapnik. Support-Vector Networks.[J]. Machine Learning,1995,20(3).
* [6] V. N. Vapnik, An overview of statistical learning theory, IEEE Trans. Neural Netw. 1999, 10, 988-999.
* [7]黄勇,郑春颖,宋忠虎.多类支持向量机算法综述[J].计算技术与自动化,2005(04):61-63.
* [8]奉国和.SVM分类核函数及参数选择比较[J].计算机工程与应用,2011,47(03):123-124+128.

主要code附录：

divideTrainTest用于分出训练集和测试集

featureExtra用于提取原始数据特征

knn用于knn聚类

main为主函数

neuralNetwork用于搭建神经网络

normalize用于归一化

plotpictures用于展示运行结果

project用于将数据进行投影

svm用于支持向量机分类

usv用于数据的奇异值分解

main.m：

clear;

%% 数据读取与预处理

%test\_rate代指测试集占整体数据集比例

%featureData为归一化后的原始图片hog+hsv特征

load('.\cifar-10-batches-mat\data\_batch\_1.mat')

test\_rate=0.1;

featureData=featureExtra(data);

[train\_data,train\_label,test\_data,test\_label]=divideTrainTest(featureData,labels,test\_rate);

%% pca降维处理

% pca\_store\_rate为pca保留特征占总特征的比例

pca\_restore\_rate=0.95;

[pca\_u,restore\_dim] = usv(train\_data,pca\_restore\_rate);

pca\_train\_data= project(train\_data,pca\_u,restore\_dim);

pca\_test\_data= project(test\_data,pca\_u,restore\_dim);

%% 神经网络分类

%net为训练后网络，nn\_accuracy为神经网络测试集正确率

[nn\_net,nn\_accuracy] = neuralNetwork(pca\_train\_data,train\_label,pca\_test\_data,test\_label);

%% svm分类

%svm\_model为训练后模型，svm\_accuracy为支持向量机测试正确率

[svm\_model,svm\_accuracy] = svm(pca\_train\_data,train\_label,pca\_test\_data,test\_label);

%% knn分类

%knn\_accuracy为最近邻算法分类正确率

knn\_accuracy = knn(pca\_train\_data,train\_label,pca\_test\_data,test\_label,5);