

模式识别作业6

张蔚桐 2015011493 自55

2017 年 4 月 18 日

1

1.1

首先由分类面形状可以知道(C)图必然为线性核函数得到的结果，同时也可以得到(D)是精细径向基($\sigma = 0.1$)得到的结果，(A)是二次多项式核得到的结果。同时，观察支持向量的位置和分类面和线性分类面的差异可以看出，(E)是三次多项式核函数的结果，而(B)是粗糙径向基得到的结果($\sigma = 1$)，(F)是中等径向基得到的结果($\sigma = 0.5$)。总结如下

1.2

观察数据点性质可以看出这些数据点还是可以线性可分的，因此选择线性核函数比较合理，因为比较简单，同时分类效果足够好，又不会产生过拟合情况。

2

选取数字7和9。训练时的正确率和测试时的正确率如下表所示
结果比较

经过重新训练神经网络模型，Logistic Regression，贝叶斯以及支持向量模型可以看出，神经网络模型的准确度优于其余三种线性模型，而其余

表 1: 不同分类面对应的核函数

a	二次核函数	b	$\sigma = 1$ 径向基
c	线性核函数	d	$\sigma = 0.1$ 径向基
e	三次核函数	f	$\sigma = 0.5$ 径向基

表 2: 不同分类面对应的核函数

核函数种类	训练时正确率	测试时正确率	调整后训练正确率	调整后测试正确率
线性核函数	96.2%	96.32%	96.51%	96.51%
二次核函数	98.2%	98.04%	98.20%	98.87%
三次核函数	98.6%	98.43%	98.90%	99.17%
精细径向基	51.2%	51.55%	99.00%	99.02%
中等径向基	97.7%	97.94%	96.40%	96.61%
粗糙径向基	96.2%	96.22%	95.30%	94.80%

三种线性模型模型（包括支持向量机采用线性核函数的时）基本相同，均在96%上下浮动，而神经网络模型可以达到98%以上。

当支持向量机采用非线性模型且没有出现明显的过拟合的情况下，其正确率相对线性模型也得到了提高，最高也可到达98%附近。同时可以看出，在没有出现过拟合的情况下时，非线性程度越高，准确性越好，非线性程度越低，和线性核函数的估计越相近。

但是当支持向量机采用精细径向基时，出现了明显的过拟合情况，即使是在训练集上也出现了明显的过拟合情况。

进一步研究发现，如果将训练过程中的“变量标准化”选项设置为false，则准确率会得到大幅度的提高，考虑其原因可以认为是图片中存在大量始终为零或在零附近的因为噪声引起的变化的特征，采用变量标准化后，程序认为每一个变量均是[0,1]之间的值，因此大大增加了这些噪声的干扰以及全零列的巨大误差，因此不进行变量标准化提高了正确率，尤其是对精确径向基而言。需要注意的是，这里的变量标准化是将每一维特征标准化，而不是将整个输入统一标准化，对整个输入统一标准化基本不会影响准确率。

在取消变量标准化之后，可以看出随着模型非线性程度的加大，系统的正确率不断提高，其中三次核函数，精细径向基和神经网络的正确率相近，而粗糙径向基，线性核函数，线性方法的正确率相近。

神经网络模型的混淆矩阵附加如图1.2所示



图 1: 神经网络在训练集上的混淆矩阵

