THU-70250043 Pattern Recognition (Spring 2018) Homework: 5

稀疏学习，距离度量，最近邻方法

Lecturer: Changshui Zhang zcs@mail.tsinghua.edu.cn
Student:

1.证明利用欧式距离的最近邻规则将空间划分成的区域（Voronoi网格）是凸的。

2.假设数据$x \in \mathbb{R}$，其类别w_i的先验概率为$P(w_i) = \frac{1}{c},\ i = 1,2,\ldots,c$，且有：

$$p(x|w_i) = \begin{cases} 1, & 0 \le x \le \dfrac{cr}{c-1} \\ 1, & i \le x \le i + 1 - \dfrac{cr}{c-1} \\ 0, & \textit{其他} \end{cases}$$

其中$0 < r < \frac{c-1}{c}$，证明：

1）贝叶斯误差率为：$P^* = r$；

2）最近邻规则的误差率等于贝叶斯误差率。

3.证明 Minkowski 距离是一个距离度量。

4.编程实现最近邻及 K 近邻方法，在 MNIST 数据集上测试，并撰写实验报告。
要求：
1）使用不同规模的训练样本，比较最近邻分类器的性能变化，包括正确率，时间和空间复杂度等；
2）使用不同的 k 值，分析对性能的影响；
3）使用不同的距离度量，分析对性能的影响；
4）是否存在一组不全为 0 的系数a_k，使得数据经过$x'_k = a_k x_k, k = 1,2,\ldots,d$的变换之后，使用最近邻分类器的效果得到提升？如果存在设计一组。
5）在最近邻分类器中，设计切线距离代替欧氏距离，叙述计算方法，并比较 MNIST 上分类器性能的变化。
MNIST 是 0-9 的手写数字图片，训练集 60000 张，测试集 10000 张。（如果计算量过大内存放不下，可以使用训练集的一部分进行实验。）