

练习题

1. 写出核感知器算法，在 TwoMoon.mat 数据集上测试核感知器算法的性能。其中一半样本用于训练，一半样本用于测试。

2. 写出 k-均值(k-means)算法的关键步骤，尝试应用核技巧，给出核化的 k-均值聚类算法；并在下列数据集上比较 k-means 算法和核化的 k-means 算法的性能。

(1) YaleB_1024x640_10class.mat; (2) BinaryDigs_320x390_10class.mat;

(3) Novartis_BPLC_zeromean_unitL2.mat (4) LungA_local_zeromean_unitL2.mat

(5) leukemia_zeromean_unitL2.mat

提示: (1) 首先需要搞清楚如何计算聚类结果的准确率(包括 Accuracy / Rand Index); (2) 选择何种 Kernel 函数。

3. 设计一个 2 层感知器，其中输入维数为 d , 隐藏层包含 m 个神经元,输出层包含 c 个神经元,激活函数选用 Logistic 函数，如果使用 MiniBatch 形式的平方误差代价函数：

(1) 画出全连接网络的结构图；

(2) 画出输出层神经元 k 的信号流图，推导神经元 k 的权值 w_{ki} 的更新公式；

(3) 画出隐藏层神经元 j 的信号流图，推导神经元 j 的权值 w_{ji} 的更新公式。

4. 设计一个 2 层感知器，其中输入维数为 d , 隐藏层包含 m 个神经元,输出层包含 c 个神经元，选用平方误差代价函数，激活函数选用 Logistic 函数。请在数据集 USPS(USPS_256x7291.mat)上训练一个 MLP(d - m - c), 其中 $d=256+1$, 输出层神经元数目 $c=10$, 隐藏层神经元数目 m 自行确定。每个类别保留 70% 样本用于训练，剩余样本用于测试。

(1) 统计各个类别训练样本和测试样本的分布，确认样本分布是否均衡；

(2) 画出训练过程中训练集上的分类错误率；

(3) 画出测试集上的分类错误率。

(4) 画出测试集上的分类错误率。

5. 设计一个 2 层感知器, 其中输入维数为 d , 隐藏层包含 m 个神经元, 输出层包含 c 个神经元, 选用平方误差代价函数, 激活函数选用 ReLU 函数。请在数据集 USPS(USPS_256x7291.mat)上训练一个 MLP(d - m - c), 其中 $d=256+1$, 输出层神经元数目 $c=10$, 隐藏层神经元数目 m 自行确定。每个类别保留 70% 样本用于训练, 剩余样本用于测试。

(1) 统计各个类别训练样本和测试样本的分布, 确认样本分布是否均衡;

(2) 画出训练过程中训练集上的分类错误率;

(3) 画出测试集上的分类错误率。

(4) 画出测试集上的分类错误率。

6. 在问题 4 中, 如果选用交叉熵定义代价函数, 如何训练? 结果如何?