

《计算机应用数学第二次作业》

截止时间：2022.12.08

计算题（40 分，每题 10 分）

1、
求解线性方程组

$$\begin{cases} x_1 - 2x_2 + 3x_3 - x_4 - x_5 = 2 \\ x_1 + x_2 - x_3 + x_4 - 2x_5 = 1 \\ 2x_1 - x_2 + x_3 - 0 - 2x_5 = 2 \\ 2x_1 + 2x_2 - 5x_3 + 2x_4 - x_5 = 5 \end{cases}$$

2、当 a,b 取何值时线性方程组

$$\begin{cases} x_1 + x_2 + x_3 + x_4 + x_5 = 1 \\ 3x_1 + 2x_2 + x_3 + x_4 - 3x_5 = a \\ x_2 + 2x_3 + 2x_4 + 6x_5 = 3 \\ 5x_1 + 4x_2 + 3x_3 + 3x_4 - x_5 = b \end{cases}$$

有解，并求解。

3. 假设有四个观测点,分别为(1,1), (2,1), (3,2)和(3,3)，用最小二乘法的两种解法（普通方法和矩阵方法）求解直线 $y=ax+b$ 使其最佳拟合这些散点。

4. 现有如下数据（4 个维度），请使用 PCA 将数据降到 2 维，给出详细计算过程。

A	B	C	D
1	5	3	1
4	2	6	3
1	4	3	2
4	4	1	1
5	5	2	3

编程题（60 分）

1、基于降维的机器学习

数据集：kddcup99_Train.zip, kddcup99_Test.zip

数据描述：<https://www.kdd.org/kdd-cup/view/kdd-cup-1999/Data>，目的是通过 42 维的数据判断某一个数据包是否为 attack。(注意，数据中有多种类型的 attack，我们将他们统一认为是 attack，不做具体区分，只区分 normal 和 attack，可以预处理的时候把所有的 attack 都统一 re-label，具体 attack 列表可参考

https://kdd.org/cupfiles/KDDCupData/1999/training_attack_types)。

任务描述：任选 1 种 Classification 模型（具体模型不限，例如贝叶斯、决策树等，除了深度学习模型，可自由选择），结合两种降维方法，PCA 与 AutoEncoder，实现数据降维+分类。

要求输出：不同方法降低到不同维度对判别结果的影响（提升还是下降），Plot 出两种降维方法降到 2 维之后的结果（Normal 和 Attack 的 sample 用不同颜色）

Optional：尝试其他 Feature Selection 方法，例如 MRMR 等。

2、深度学习训练方法总结

数据集：kddcup99.zip

任务描述：实现一个简单的神经网络模型判别数据包是否为 attack，网络层数不小于 5 层，平台不限（TensorFlow、Pytorch 等都可），例如 42->36->24->12->6->1。尝试至少 2 种激活函数，至少 2 种 parameter initialization 方法，至少 2 种训练方法 (SGD, SGD+Momentum, Adam)，训练模型并判断训练结果

要求输出：1) 模型描述，层数，每一层参数，激活函数选择，loss 函数设置等；2) 针对不同方法组合（至少 4 个组合），plot 出随着 epoch 增长 training error 的 test error 的变化情况。