

## 一、开放性实验

### 1. 数据集

#### ①Adult Dataset

Predict whether income exceeds \$50K/yr based on census data.

<http://archive.ics.uci.edu/ml/datasets/Adult>

adult.data为训练集，adult.test为测试集。

数据预处理代码可参考：

<https://github.com/saravrajavelu/Adult-Income-Analysis>

#### ②Bank Marketing Dataset

Predict if the client will subscribe a term deposit.

<http://archive.ics.uci.edu/ml/datasets/Bank+Marketing>

使用bank-additional.zip文件。

随机使用75%的数据做为训练集，其余为测试集。

数据预处理代码可参考：

<https://github.com/juliencohensolal/BankMarketing>

### 2. 实验内容

- (1) 数据集选择：在以上两个数据集中任意挑选一个感兴趣的数据集。
- (2) 数据预处理：数据集预处理可参考相关项目，不做硬性要求。
- (3) 模型：任意选择一个本门课程接触到的神经网络进行以上**分类**任务（逻辑回归、CNN、RNN...）
- (4) 回答以下问题
  - ① 模型有没有出现过拟合现象？引入任意一种正则化方法（如L2正则化、Dropout等）对结果是否有提升？

② 不同的优化算法对结果是否有影响？引入任意一种其他的优化算法进行比较。（如SGD、Adam等）

### 3. 评分说明

分类结果不作为主要的评判标准，侧重对模型设计、实验问题的分析和说明。  
对模型结构或重要超参数的选择进行进一步探索为加分项。

## 二、实验注意事项

1. 本作业不要求使用GPU环境进行实验，初步考虑第十周会继续开放实验室，有兴趣的同学可以来上机实验。

2. 作业提交时间：

第三次作业：11月11日

第四次作业：11月23日

## 三、考试相关

在实验上，本学期我们尝试丰富了这门课的实验内容，实验难度有所提高。相对于往年，期末考试可能会有个别的计算题或推导题，总体更侧重于对课程中重点概念与内容的理解，难度会适当降低。感谢同学们本学期的配合与支持，祝同学们取得好成绩！

有任何问题和建议欢迎大家和助教联系：937332325@qq.com