**2018机器学习大作业题目**

不限于以下题目，可以自行拟定题目，例如参考国外机器学习案例思路（参考思路，不可照搬数据和源代码），一定要有自己的工作量。最终每个题目都要提交工程报告、训练数据、工程源代码（github）、打包程序。以个人或小组为单位，如果以组为单位，每个人都要提交工程报告，围绕自己的工作写工程报告，需要由组长分配好每人的工作量（按照百分比），并在提交时向老师说明，不可均分。最终大作业分数会根据题目本身难度、题目完成程度、工作量、工程报告结构和内容、界面展示度等决定。

1.收集一定数量的学生成绩数据（例如10个男生，10个女生，10门课），根据成绩数据，预测该学生是男生还是女生。（分类问题，可以使用libsvm或者神经网络软件包）（**邓斌组，刘鹏飞组**）

提交：工程报告+训练数据+工程源代码+打包程序

2.针对合肥某个特定小区，在合肥二手房网站上（例如http://fangjia.fang.com/hf/）， 收集这个小区历史的房屋月均价数据（例如从去年5月到今年3月），预测下个月（如今年4月）的房价。（单变量线性回归问题）（**王神亚组，房飞龙和马彦辰**）

提交：工程报告+训练数据+工程源代码+打包程序

3.在libsvm的数据集合中，选取一个多分类数据集合，利用不同的libsvm核函数、和参数进行训练libsvm，并比较实验结果。至少要验证3个以上的核函数，并尝试5组以上的参数（实验）。https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/。（libsvm工具包的使用）

提交：工程报告+训练数据+工程源代码+打包程序

4.基于scikit-learn软件包，对mnist数据集合进行分类，并可视化。至少要尝试5种不同的分类方法，并比较分类结果。（王健帅组）

提交：工程报告+训练数据+工程源代码+打包程序

5.收集一定量的猫和狗的图片（例如100个猫的图片，100个狗的图片），训练一个分分类器，能够判断输入的图片中是猫还是狗。图像处理和分类器可以直接使用工具包构建。（王贯东个人，单泽彬个人）

提交：工程报告+训练数据+工程源代码+打包程序

6. 从http://archive.ics.uci.edu/ml/上下载任意一个数据集，然后根据其问题类型，用相应的软件包，构建相应的分类器或者回归模型。（李文琪个人，）

提交：工程报告+训练数据+工程源代码+打包程序

7.收集合肥地区过去一段时间（例如过去一年每个月的平均值）的空气质量（例如pm2.5值），然后构建回归模型，能够预测今年某个月的空气质量值。

提交：工程报告+训练数据+工程源代码+打包程序

8.收集一定数量的男和女的照片，训练一个分类器，能够直接根据照片分类出男女。（王鑫组，阚雨露个人，）

提交：工程报告+训练数据+工程源代码+打包程序

9.收集一定数量的男生的手写字（图片）和一定数量的女生手写字（图片），然后训练一个分类器，能够直接根据这个手写字的照片分类出男女。（李星达组，王嘉辉个人）

提交：工程报告+训练数据+工程源代码+打包程序

**10.基于深度学习的面部表情分类**

自行查找关于深度学习的资料，基于tensorflow等深度学习系统，构建一个面部表情分类，可以检测静态图片中的人脸面部表情，或者面部表情实时监测系统。（李佳组，曹天恒和蒙枭）

提交：工程报告+训练数据+工程源代码+打包程序

**11.新闻文本分类**

从新闻网站上，收集各类新闻文本（5-10种），根据课堂上讲解的one-hot特征提取方式，或者自行查找文本特征提取方式，对文本进行提取特征并向量化，利用机器学习模型进行分类，构建一个新闻文本分类系统。（丁中华个人）

提交：工程报告+训练数据+工程源代码+打包程序

12**.垃圾邮件分类系统**

从邮件系统中，收集垃圾邮件和非垃圾邮件每类50-100组，根据课堂上讲解的垃圾邮件检测技术，进行垃圾邮件分类，构建垃圾邮件分类器。（江娟焦佳敏组队，陈凯琪）

提交：工程报告+训练数据+工程源代码+打包程序

**13.股票价格预测**

从股票网站上，收集某股票的历年价格，对历年每月（或每天）价格数据进行建模，预测其下月（或下一天）的股票价格。考虑股票价格和时间的关联。（谢丕捷个人）

提交：工程报告+训练数据+工程源代码+打包程序

**14.饭卡消费预测系统**

收集自己或同学的饭卡消费数据，从校园网账户中，对每月（每天）的消费数据进行建模，预测下月（或者下一天）的饭卡消费。考虑饭卡消费和月份之间的关联。

提交：工程报告+训练数据+工程源代码+打包程序

**15.水果检测**

设计一个水果检测系统，可以对3-4种水果进行分类，收集3-4种水果的照片，每种水果收集20-50张照片，用机器学习模型进行学习训练，构建桌面应用或者移动端应用，推荐开发移动端应用，实现水果的实时检测与分类。（张方勇，许文婕组（王笑宇、张煜））

提交：工程报告+训练数据+工程源代码+打包程序

**16.国籍检测**

设计一个小型的国籍分类系统，能够分类出中国人和日本人。通过摄像头，或者搜索引擎，搜索50-100张日本人和中国人头像照片，训练神经网络（可以直接用课程中讲的神经网络模型），或者通过SIFT特征提取+训练SVM，或者其他任何机器学习模型，完成桌面应用或者移动应用开发。课时实现对一张静态照片中的人脸进行国际检测，或者进行实时的国籍检测。（邓斌组，张方勇个人）

提交：工程报告+训练数据+工程源代码+打包程序

**17. 基于机器学习的图片自动拼接**

自行查找相关模型算法，构建一个能够将两个图片进行自动拼接为一个完整图片的系统，类似于全景照片的拼接过程。

提交：工程报告+训练数据+工程源代码+打包程序

18.自行设计的解决实际问题的题目，或者参加国际国内比赛的机器学习相关项目。

提交：工程报告+训练数据+工程源代码+打包程序

Titanic 申世浩，邹安宁，马宽，余鸿辰

衣服类型分类 王梓静组

张杰林 2016213593 房价预测

王远见 多变量线性回归的，用在卫星图像融合