多标签学习 （100分）

（multi-label learning）

（一）基础要求-Baseline（40分）：

（1） 回归模型+正则化项（）实现多标签分类任务，测试三种正则化项对模型性能的影响，解释其原因；

（2） 运用不同的损失函数Hinge Loss、Logistic Loss，对比性能并分析原因；

（二）中级要求（30分）：

（1）二选一即可

* 实现有监督情况下的大规模多标签分类，具体的方法见参考文献[1][2][3]，对比Baseline分析性能提升原因；
* 自选方法实现本课题-有监督大规模多标签分类问题，对比Baseline分析性能提升原因；

（三）高级要求（30）

注：在高级要求中，假设：标签有缺失值，missing labels or incomplete data问题；

（1）二选一即可

* 在参考文献中[2-6]任选一种方法实现，且保证性能较（二）有所提升；
* 自选方法实现本课题-标签不完整情况下的多标签分类任务；

（2）解释性能提升原因；

（3）尝试结合特征选择方法，实现多标签特征选择和多标签分类，并分析多标签学习性能；

（四）作业格式

作业提交格式包括：课堂展示（40分）+作业文档（60分）

（1）每组需要一名同学**课堂展示**课程设计，需做PPT或者PDF，每组8-10分钟，答辩时间另行通知；

（2）每组需提交**全部源码、课堂展示PPT（PDF）**和**设计报告**,设计报告内容如下：

* 问题提出
* 主要设计思路及处理方法
* 实验设计及说明
* 实验结果及分析
* *有关信息：小组成员学号，姓名，联系方式*

（3）文件提交方式同之前作业提交格式（邮箱），**每组只需提交一份作业**

命名方式举例：第一组+王子义+任金凯+许越凡.rar

（4）报告提交时间：**2019.6.18**

Reference

[1] Large-scale multi-label learning with missing labels, ICML’2014.

[2] Sparse local embeddings for extreme multi-label classification, NIPS’2015.

[3] Robust Extreme Multi-label Learning, KDD’2016.

[4] Image tag refinement towards low rank content tag prior and error sparsity. Multimedia’2010.

[5] Learning from Weak and Noisy Labels for Semantic Segmentation. TPAMI’2017.

[6] SemiBoost: Boostting for semi-supervised learning. TPAMI’2009.

注：

（1）数据集和参考文献见Dataset和Ref文件夹；

（2）本课题的评价指标统一使用准确率（precision）和召回率(recall)；

（3）不建议使用CNN、RNN、LSTM等深度学习模型；

（4）1-3人一组，每组需一人展示成果（时间另行通知），请在报告以及报告展示中写清任务分工；

（5）论文检索关键词：multi-label learning/feature selection、missing labels、incomplete labels/data。

（5）设计报告中不要粘贴源代码；

如有问题欢迎随时与我们联系：[xuyuanyuan@mail.nankai.edu.cn](mailto:xuyuanyuan@mail.nankai.edu.cn)、yuyin@mail.nankai.edu.cn