

# 第 1 章 Final Project

本课程 Final project 的目的旨在激发学生在课程讲授内容之外, 对本学科领域进行主动探索, 阅读, 翻译相关文献, 对某一个具体专题进行探索, 调研和总结; 或者将所学知识应用于相关科研/应用问题的建模和求解过程中, 增强解决实践中具体问题的能努力. 因此, 本课程 FP 可以自由选择下面论文阅读1.1.1, 材料阅读与翻译1.1.2或其他问题1.1.3之一.

## 1.1 作业内容

任意选择以下文章之一, 通过阅读, 理解, 分析, 和复现论文实验结果. 可以借助作者程序或者第三方程序, 并鼓励自行实现.

### 1.1.1 论文阅读

注: 点击对应的问题, 会直接跳转相应的论文 pdf 页面.

#### 1. Approximate Newton Methods, JMLR, 2021.

二阶优化算法的主要困难在于计算和存储 Hessian 矩阵造成较大的成本, 使得算法难以扩展到维度较高的问题上. 对于机器学习问题, 借助随机梯度近似和矩阵近似, 可以避免较高的复杂度.

#### 2. Second-order stochastic optimization for machine learning in linear time, JMLR, 2017.

本论文通过矩阵近似等方法, 他出一种计算效率线性的二阶优化方法.

#### 3. Accelerating stochastic gradient descent using predictive variance reduction, NIPS, 2013.

随机梯度下降 (SGD) 算法是深度学习能够获得广泛应用的一个重要保证. 一般地, 由于 SGD 的 mini-batch 具有方差, 使得算法的收敛性受到方差的影响. 方差缩减技术, 通过对单个样本梯度的噪声的复用, 有效降低梯度估计中的噪声, 提高算法收敛效率.

#### 4. Accelerating Quadratic Optimization with Reinforcement Learning, NIPS, 2021.

近年来, 使用强化学习直接学得某一类问题上的优化算法称为一个热点问题, 针对此问题展开阅读/翻译和调研. 此问题适合对强化学习感兴趣并有所了解的同学.

#### 5. Identifying and attacking the saddle point problem in high-dimensional non-convex optimization, NIPS, 2014.

深度神经网络的 landscape 上, 大部分临界点是鞍点而不是局部极小点. 本论文使用一种线性信赖域方法验证此性质.

## 6. How Does Batch Normalization Help Optimization?, NIPS, 2018.

批归一化 (Batch normalization) 是一种深度学习中常用的技巧, 通过在层内部对数据进行归一化处理降低数据的内部偏移, 从而加速算法收敛. 本论文从优化的角度来分析, BN 如何有效提升了优化效率.

## 7. Natural Evolution Strategies, JMLR, 2014

本论文提出一种将无导数随机优化方法, 通过将原问题转化为将概率分布上的优化问题, 并通过概率分布上的自然梯度下降-采样估计梯度的方法, 获得一大类的优化方法.

## 8. Efficient decision-based black-box adversarial attacks on face recognition, CVPR, 2019

机器学习模型, 尤其是深度学习模型, 通常对于输入数据较为敏感. 当输入图像有微小的人眼不可识别的扰动时, 可能会对深度神经网络模型的输出造成重大影响, 使得分类出现错误. 这样的扰动样本称为对抗样本. 如何高效生成高质量对抗样本, 对于提升模型鲁棒性具有重要意义. 将输入图像的扰动作为变量, DNN 的输出作为目标函数, 可以将上述问题建模为优化问题, 并使用相应算法求解. 本问题针对深度学习模型的对抗样本生成问题开展调研, 分析.

## 9. Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines

## 10. The landscape of deep neural networks and saddle points

## 11. 其他问题, 后续会继续增加.

## 1.1.2 材料阅读与翻译

选择以下材料之一阅读并翻译.

注意, 选择翻译材料的同学, 应完成整个材料. 如果选择此部分内容, 每位同学只可以选择对应学号尾号的材料进行翻译, 以避免所有人都选同一材料. 例如, 学号尾号为 7, 那么  $7 \bmod 4 = 3$ , 应选择 M3 文档. 如果两人组队, 任选一位同学学号尾号即可.

## 1.1.3 其他问题

自行选择其他最优化论文, 阅读分析, 写成研究报告.

注: 如果自行选择论文, 请尽快和导师或者助教同学联系确定是否可行. 论文应该以优化为主题 (包括优化算法或者建模).

## 1.2 要求

- 最终形成书面报告, 报告应使用本模板, 篇幅 5 页以上, 不包含参考文献.

- 提交电子版报告, 并提交相应 L<sup>A</sup>T<sub>E</sub>X 源码.
- 关于组队: 可以两人组队, 报告应注明每个人所完成内容和贡献比例, 组队的书面报告应达到 8 页以上, 不包含参考文献.
- 关于抄袭: 禁止抄袭行为, 包括对已正式发表的中英文文献的抄袭和对一般网络文章, 博客, 公众号介绍内容的抄袭. 但是允许相互讨论.

## 1.3 报告的形式

选择以下论文之一, 阅读, 理解, 分析, 总结, 复现相应实验结果, 最终完成报告. 报告应包含以下内容:

1. 文章要处理的问题是什么, 这个问题为什么重要?
2. 此问题存在什么难点? 现有的其他人/其他方法是怎样处理这些难点的? 这些方法有什么优缺点?
3. 本文使用什么方法来处理问题的难点? 这样处理有什么优点?
4. 实验效果如何? 从实验效果上来说, 相对其他方法有什么优缺点?

## 1.4 关于 L<sup>A</sup>T<sub>E</sub>X 格式的提示

1. 公式采用 ‘equation’ 环境, 可以通过 ‘label’、‘eqref’ 命令产生交叉引用;
2. 图片、表格居中, 格式正确 (如 caption 位置等), 通过 ‘label’、‘ref’ 命令做好交叉引用;
3. 图片尽可能采用 Acrobat 裁剪的矢量图;
4. 做好参考文献的交叉引用;
5. 公式末尾添加对应的标点符号.
6. 采用 TikZ 重新绘制流程图、示意图 (原书没有可以自己加);
7. 对于对应算法给出代码实现, 并上传到 GitHub 仓库, 在作业中做好链接;
8. 采用 biblatex 包完成参考文献的反向索引 (backref);
9. 根据章节内容出 1-3 道有意思的习题, 并收集自己及不少于 3 个同学 (如舍友等) 的解答;
10. 整体完成美观。