实验用数据集

* 图像分类
  + MNIST
    - Superloss文章中使用，但作为回归任务是实现的，所以baseline用的是我们自己实现的baseLine
    - 下表初步结果（metric:accuracy）：括号数字表示最好的结果出现的epochs个数，其他超参文中有说明
      * B1:no CL(还没做，需要跑下出一组结果，需要5 runs)
      * B2:Superloss（我们自己跑出来的，需要5 runs）
      * 调参方向：ours好于B2（可能好于）B1
      * Ours:需要5 runs

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| GPU24 | Tranh参数 | **Our+tranh** | **Superloss** | **Baseline** |
| Noise=0 | Wd=1e-6 | **98.95（16）** | **98.92（6）** |  |
| Noise=20 | Wd=1e-6 | **98.83（19）** | **98.8（19）** |  |
| Noise=40 | Wd=1e-6 | **98.49（10）** | **98.41（14）** |  |
| Noise=60 | Wd=1e-6 | **97.89（12）** | **97.68（5）** |  |
| Noise=80 | Wd=1e-6 | **95.68（15）** | **95.8（14）** |  |
|  |  |  |  |  |
| GPU24 |  | **Our+tranh** | **Superloss** | **Bas****eline** |
| Noise=0 |  |  |  |  |
| Noise=20 |  |  |  |  |
| Noise=40 |  |  |  |  |
| Noise=60 |  |  |  |  |
| Noise=80 |  |  |  |  |

* + CIFAR10
    - Baseline直接用原文结果
      * B1:Data parameters
      * B2:DivideMix
      * B3:Superloss
      * B4:
      * Ours:需要5 runs

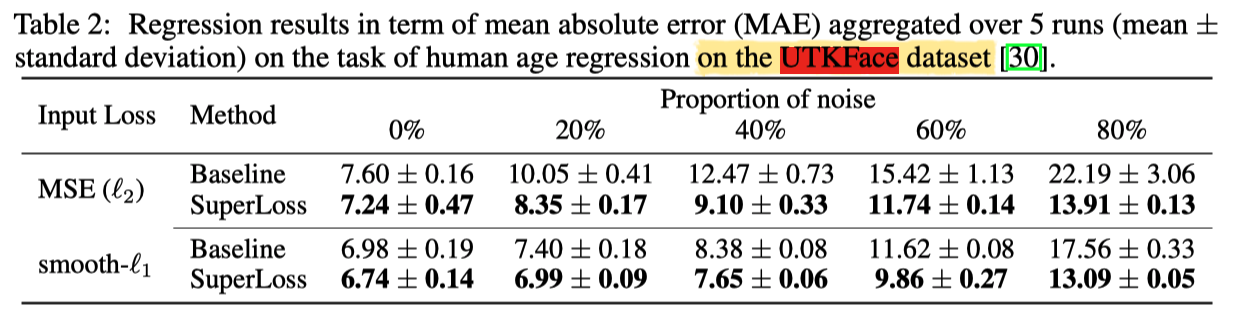
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| GPU24 | Tranh参数 | **Our+tranh** | Data para | Superloss | Baseline |
| Noise=0 | Wd=1e-6 | **95.13(117)** | - | - |  |
| Noise=20 | Wd=1e-6 | **94.04(107)** | - | 93.390.12 |  |
| Noise=40 | Wd=1e-6 | **92.26(104)** | 91.10 | 91.730.17 |  |
| Noise=60 | Wd=1e-6 | **86.65(100)** | - | 85.480.13 |  |
| Noise=80 | Wd=1e-6 | **68.72(82)** |  | 77.490.29 |  |

* + CIFAR100
    - Baseline直接用原文结果
      * B1:Data parameters
      * B2:DivideMix
      * B3:Superloss
      * B4
      * Ours:需要5 runs

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| GPU24 | Tranh参数 | **Our+tranh** | Data para | Superloss | Our\_ea | Baseline |
| Noise=0 | Wd=1e-6 | **80.29（111）** | 79.6(111)/自己报80.80.1 | 78.85(112) | 78.77(111) | Data parameter table 1:80.10.2自己报 |
| Wd=0.1  曲线看不到 | **80.63（116）** |
| Noise=20 | Wd=1e-6 | **75.36（108）** | 75.680.12 | 74.73 | 75.72（113） |  |
| Noise=40 | Wd=1e-6 | **71.15（81）** | 71(84,early stop)自己报70.93%0.15 | 71.05% 0.08(avg) | 71.19(103) |  |
| Wd=0.1 | **70.1（81）** |
| Noise=60 | Wd=1e-6 | **60.02（85）** | - | 62.390.17(expavg) | 58.96（103） |  |
| Noise=80 | Wd=1e-6 | **18.3（118）** | 35.8 | 37.910.93(logC) | 9.52（106） |  |

* 图像回归
  + UTKFACE
    - Baseline直接用原文结果
      * B1:no CL(见原文附录)
      * B2:Superloss(见原文附录)
      * Ours:需要5 runs（其中，L1 loss，noise=60情况已完成调参，结果如下）
        + best mae: 9.677920588859774, params: {'lr': 0.1, 'lr\_inst\_param': 0.0008292186575905661, 'sup\_eps': 3.0, 'tanh\_a': 0.2363806481804235, 'tanh\_p': 1.6890393059052589, 'tanh\_q': -65.0, 'wd\_inst\_param': 0.0006204837469530673}, iteration: 7, status: ok

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| GPU24 | Tranh参数 | Our+emak\_tanh(两阶段) | |  |  |
| **L1** | **L2** |  |  |
| Noise=0 | Wd=1e-6 | **5.42（98）** | **5.40（98）** |  |  |
| Noise=20 | Wd=1e-6 | **5.75（91）** | **6.78（75）** |  |  |
| Noise=40 | Wd=1e-6 | **6.91（63）** | **8.62（66）** |  |  |
| Noise=60 | Wd=1e-6 | **9.68**  **【已完成调参】** | **10.70（44）** |  |  |
| Noise=80 | Wd=1e-6 | **13.08（96）** | **12.86（35）** |  |  |



* NLP分类
  + WIKIHOW
    - 有baseline，如果不超过不强求，可用我们自己的baseline【原文中已有的超参，我们是否和它一样？】
      * B1:no CL需要5 runs
      * Ours需要5 runs
      * 出表格结果：noise=0/20/40/60/80%对应的accuracy
      * Pretrained language model不限，roberta,bert都可
* NLP回归
  + Digitsum
    - 【没有baseline?】如果没有，用我们自己的
      * B1:no CL需要5 runs
      * B2:baby step需要5 runs
      * Ours需要5 runs
      * 当前的loss和metric是什么？
      * 出表格结果：noise=0/20/40/60/80%对应的metric

|  |  |  |  |
| --- | --- | --- | --- |
| GPU24 | **Our+tranh** | **Babystep** | **Baseline** |
| Noise=0 |  |  |  |
| Noise=20 |  |  |  |
| Noise=40 |  |  |  |
| Noise=60 |  |  |  |
| Noise=80 |  |  |  |

注意

* 同一dataset，不同noise比例可能采用不同超参数
* 在完成调参后，进行5runs
* 除更新相应表格，请附上对应的超参数取值，e.g., DigitSum中使用的LSTM的一些超参数等，便于文章实验部分描述，如果和原文超参数一直的话，请描述下哪些相同，哪些不同（对应的取值）
* Dicrimloss中的k同样有三种取值方法，即超参数直接指定，全局平均和滑动平均，报最好结果