# Update on Varying Learning Rates for 3 Models on CIFAR-10

Gobind Puniani

MATH 198: Spring 2020

01/30/20

# Semester Goals

- Present a research poster with meaningful results at a conference
  - AMS Sectional Meeting
- Be able to explain neural networks and activation functions to any competent undergraduate student in under 3 minutes
- Read and summarize 2 papers per month
- Be able to write Python scripts without Patrick's help

# TLALOC Runs

- 3 models:
  - DenseNet-121
  - MobileNet v2
  - SE Net-18
- 5 Learning Rates (factors of 10):
  - $10^{-4}$ to $10^{0}$
- 5 activation functions:
  - ReLU
  - Swish
  - Mish
  - TAct
  - mTAct
- 3 runs each (averaged)

# DenseNet 121 Test Top-1 and Top-3 Accuracy

| LR | Mish | ReLU | Swish | TAct | mTAct |
|---|---|---|---|---|---|
| 0.0001 | 0.829608 | 0.369891 | 0.880274 | 0.369495 | 0.370451 |
| 0.0010 | **0.904964** | **0.910733** | **0.908788** | **0.907700** | **0.908986** |
| 0.0100 | 0.815104 | 0.365605 | 0.889933 | 0.361353 | 0.362111 |
| 0.1000 | 0.813753 | 0.371671 | 0.888581 | 0.371242 | 0.368110 |
| 1.0000 | 0.558478 | 0.100508 | 0.850442 | 0.099585 | 0.100046 |

| LR | Mish | ReLU | Swish | TAct | mTAct |
|---|---|---|---|---|---|
| 0.0001 | 0.967464 | 0.528745 | 0.968717 | 0.528844 | 0.529964 |
| 0.0010 | **0.985924** | **0.987507** | **0.986452** | **0.986880** | **0.987276** |
| 0.0100 | 0.970069 | 0.526734 | 0.974387 | 0.526503 | 0.526932 |
| 0.1000 | 0.971189 | 0.528712 | 0.974848 | 0.530063 | 0.528217 |
| 1.0000 | 0.763680 | 0.299446 | 0.954509 | 0.299677 | 0.298292 |

# MobileNet v2 Test Top-1 and Top-3 Accuracy

| LR | Mish | ReLU | Swish | TAct | mTAct |
|---|---|---|---|---|---|
| 0.0001 | 0.315533 | 0.306863 | 0.308379 | 0.322785 | 0.326081 |
| 0.0010 | **0.863199** | **0.859935** | **0.866627** | **0.867880** | **0.852420** |
| 0.0100 | 0.361518 | 0.343289 | 0.361946 | 0.358188 | 0.356243 |
| 0.1000 | 0.113990 | 0.128329 | 0.214234 | 0.099749 | 0.083103 |
| 1.0000 | 0.100903 | 0.099815 | 0.100046 | 0.099189 | 0.099749 |

| LR | Mish | ReLU | Swish | TAct | mTAct |
|---|---|---|---|---|---|
| 0.0001 | 0.510647 | 0.508933 | 0.506560 | 0.513647 | 0.516482 |
| 0.0010 | **0.974486** | **0.975277** | **0.975541** | **0.976200** | **0.971585** |
| 0.0100 | 0.525844 | 0.522515 | 0.534283 | 0.522811 | 0.525316 |
| 0.1000 | 0.365276 | 0.325883 | 0.278844 | 0.297666 | 0.200620 |
| 1.0000 | 0.297897 | 0.298292 | 0.299117 | 0.299018 | 0.298556 |

# SE Net-18 Test Top-1 and Top-3 Accuracy

| LR | Mish | ReLU | Swish | TAct | mTAct |
|---|---|---|---|---|---|
| 0.0001 | 0.362144 | 0.877472 | 0.795425 | 0.359836 | 0.458333 |
| 0.0010 | **0.904964** | **0.901833** | **0.905558** | **0.889076** | **0.880142** |
| 0.0100 | 0.402624 | 0.888548 | 0.797304 | 0.391581 | 0.500494 |
| 0.1000 | 0.427380 | 0.874044 | 0.811116 | 0.452466 | 0.495583 |
| 1.0000 | 0.099815 | 0.854661 | 0.641054 | 0.098596 | 0.227782 |

| LR | Mish | ReLU | Swish | TAct | mTAct |
|---|---|---|---|---|---|
| 0.0001 | 0.537019 | 0.969838 | 0.919699 | 0.526206 | 0.686445 |
| 0.0010 | **0.985858** | **0.984375** | **0.985727** | **0.980914** | **0.978672** |
| 0.0100 | 0.625692 | 0.973629 | 0.921875 | 0.586300 | 0.765691 |
| 0.1000 | 0.644053 | 0.971980 | 0.933940 | 0.679655 | 0.770899 |
| 1.0000 | 0.300138 | 0.961828 | 0.810390 | 0.298062 | 0.536623 |

# Conclusion

- The original learning rate (0.0010) was optimal for all activation functions on all 3 models
- High LRs sometimes achieved high accuracy, but the average was still low due to low accuracy on 1 or 2 runs
- Low LRs simply did not have enough time to achieve high accuracy over 100 epochs