Results on Varying Learning Rates

Gobind Puniani
MATH 198: Spring 2020
02/13/20

TLALOC Runs

- 3 models:
  - DenseNet-121
  - MobileNet v2
  - SE Net-18
- 5 Learning Rates (factors of 10):
  - 10^{-4}
  - 10^{-3}
  - 10^{-2}
  - 10^{-1}
  - 10^{0}
- 5 activation functions:
  - ReLU
  - Swish
  - Mish
  - TAct
  - mTAct
- 3 runs each (averaged)
- Best in column: italicized
- Best in row: bolded

DenseNet 121 Test Top-1 Accuracy

LR	Mish	ReLU	Swish	TAct	mTAct
0.0001	0.903349	0.900481	0.904437	0.904635	**0.906909**
0.0010	*0.904964*	***0.910733***	*0.908788*	*0.9077*	*0.908986*
0.0100	0.882021	0.877637	0.882351	**0.885153**	0.877703
0.1000	0.862902	**0.885614**	0.867220	0.874571	0.885450
1.0000	0.250494	0.099848	**0.31194**	0.100046	0.100178

DenseNet 121 Test Top-3 Accuracy

LR	Mish	ReLU	Swish	TAct	mTAct
0.0001	***0.986056***	0.984342	0.985265	0.985825	0.985661
0.0010	0.985924	***0.987507***	*0.986452*	*0.986880*	*0.987276*
0.0100	0.979397	0.977749	0.980749	**0.981145**	0.980024
0.1000	0.974585	0.980353	0.977024	0.978277	**0.981309**
1.0000	0.499637	0.299941	**0.507384**	0.299215	0.299875

MobileNet v2 Test Top-1 Accuracy

LR	Mish	ReLU	Swish	TAct	mTAct
0.0001	0.738133	0.718585	0.725541	0.770866	**0.776174**
0.0010	0.863199	0.859935	0.866627	***0.867880***	0.852420
0.0100	***0.880505***	*0.862342*	*0.879747*	0.866858	*0.862540*
0.1000	0.178995	0.157404	**0.260252**	0.122594	0.098233
1.0000	**0.101925**	0.099585	0.099288	0.099914	0.099255

MobileNet v2 Test Top-3 Accuracy

LR	Mish	ReLU	Swish	TAct	mTAct
0.0001	0.929688	0.923259	0.923919	0.943730	0.947191
0.0010	0.974486	0.975277	0.975541	**0.976200**	0.971585
0.0100	***0.979727***	*0.976431*	*0.978540*	*0.977255*	*0.974321*
0.1000	0.415084	0.440961	**0.476266**	0.324862	0.222805
1.0000	0.299974	**0.300831**	0.298919	0.298820	0.297139

SE Net-18 Test Top-1 Accuracy

LR	Mish	ReLU	Swish	TAct	mTAct
0.0001	**0.882747**	0.874637	0.881659	0.876846	0.874604
0.0010	*0.904964*	*0.901833*	***0.905558***	*0.889076*	*0.880142*
0.0100	0.883801	0.886142	**0.887889**	0.879219	0.878165
0.1000	0.861650	**0.868473**	0.829608	0.832311	0.710311
1.0000	0.099585	0.100508	0.099420	0.098563	**0.157470**

SE Net-18 Top Test-3 Accuracy

LR	Mish	ReLU	Swish	TAct	mTAct
0.0001	**0.980320**	0.977683	0.978573	0.977453	0.976925
0.0010	***0.985858***	*0.984375*	*0.985727*	*0.980914*	0.978672
0.0100	0.980353	**0.981705**	0.979694	0.978771	*0.979661*
0.1000	0.972211	**0.973662**	0.964662	0.963739	0.867979
1.0000	0.299875	0.301325	0.300666	0.310160	**0.427314**

Conclusion

- The original learning rate (0.0010) was optimal for all activation functions on DenseNet-121
  - Except Mish top-3
- There was a fair amount of variability among learning rates for best performances on SE Net-18 and MobileNet v2
- The highest learning rate (1.0000) was the worst for all activation functions