

Model	Hidden layer	Optimizer	Optimal hyper parameter	Cut-off Epoch	Training		Validation		Testing	
					Averaged loss	Averaged accuracy	Averaged loss	Averaged accuracy	Loss	Accuracy
Linear classifier	\	SGD	lr=0.01, mo=0.9	16	0.378±0.002	0.827±0.001	0.381±0.014	0.827±0.006	\	\
	\	Adam	Default	16	0.375±0.003	0.829±0.001	0.378±0.015	0.828±0.005	0.349	0.850
Neural network	1 layer:	SGD	lr=0.01, mo=0.9	90	0.280±0.002	0.883±0.002	0.327±0.018	0.863±0.008	\	\
	40 units	Adam	Default	32	0.276±0.003	0.885±0.002	0.324±0.015	0.859±0.006	0.292	0.874
	2 layers:	SGD	lr=0.01, mo=0.5	128	0.261±0.003	0.891±0.002	0.398±0.060	0.829±0.026	\	\
	80, 20	Adam	Default	20	0.249±0.007	0.897±0.003	0.321±0.018	0.866±0.006	0.305	0.870
	3 layers:	SGD	lr=0.01, mo=0.9	50	0.241±0.010	0.901±0.005	0.322±0.021	0.864±0.012	\	\
	80, 40, 20	Adam	Default	18	0.230±0.008	0.906±0.004	0.316±0.017	0.870±0.007	0.316	0.872