

Table 4: Performance of ERM with weight decay (*i.e.*, L2 regularization) for linear and neural network models.

(a) Linear					(b) NN				
Dataset	L2 Reg. ( $\frac{1}{m}$ )		L2 Reg. ( $\frac{1}{\sqrt{m}}$ )		Dataset	L2 Reg. ( $\frac{1}{m}$ )		L2 Reg. ( $\frac{1}{\sqrt{m}}$ )	
	$\mathfrak{R}_{\mathcal{S}}(h)$	$\mathfrak{R}_{\mu}(h)$	$\mathfrak{R}_{\mathcal{S}}(h)$	$\mathfrak{R}_{\mu}(h)$		$\mathfrak{R}_{\mathcal{S}}(h)$	$\mathfrak{R}_{\mu}(h)$	$\mathfrak{R}_{\mathcal{S}}(h)$	$\mathfrak{R}_{\mu}(h)$
ADULT	0.207	0.207	0.248	0.248	ADULT	0.207	0.207	0.248	0.248
FASHIONMNIST	0.141	0.149	0.127	0.150	FASHIONMNIST	0.160	0.166	0.159	0.164
LETTER	0.285	0.295	0.285	0.296	LETTER	0.261	0.275	0.256	0.269
MNIST	0.067	0.092	0.066	0.092	MNIST	0.116	0.125	0.084	0.095
MUSHROOMS	0.001	0.001	0.000	0.000	MUSHROOMS	0.000	0.000	0.000	0.000
NURSERY	0.788	0.799	0.796	0.804	NURSERY	0.704	0.721	0.770	0.788
PENDIGITS	0.049	0.060	0.047	0.057	PENDIGITS	0.009	0.022	0.012	0.026
PHISHING	0.063	0.065	0.057	0.062	PHISHING	0.042	0.050	0.054	0.059
SATIMAGE	0.144	0.203	0.138	0.200	SATIMAGE	0.150	0.215	0.143	0.205
SEGMENTATION	0.058	0.157	0.075	0.177	SEGMENTATION	0.141	0.216	0.198	0.371
SENSORLESS	0.907	0.911	0.907	0.911	SENSORLESS	0.907	0.911	0.907	0.911
TICTACTOE	0.013	0.021	0.013	0.021	TICTACTOE	0.000	0.046	0.000	0.021
YEAST	0.702	0.720	0.693	0.687	YEAST	0.662	0.674	0.693	0.683