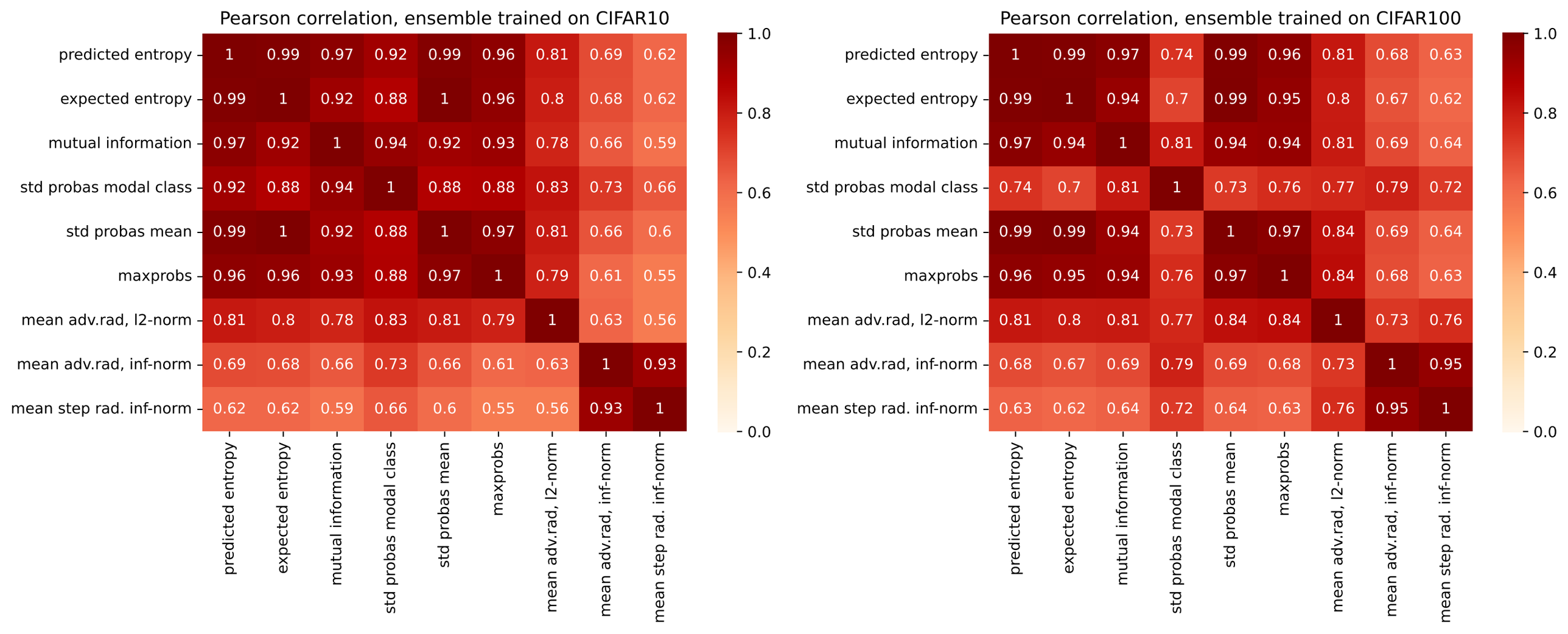
Pearson correlation, ensemble trained on CIFAR10
predicted entropy 1 0.99 0.97 0.92 0.99 0.96 0.81 0.69 0.62
expected entropy 0.99 1 0.92 0.88 1 0.96 0.8 0.68 0.62
mutual information 0.97 0.92 1 0.94 0.92 0.93 0.78 0.66 0.59
std probas modal class 0.92 0.88 0.94 1 0.88 0.88 0.83 0.73 0.66
std probas mean 0.99 1 0.92 0.88 1 0.97 0.81 0.66 0.6
maxprobs 0.96 0.96 0.93 0.88 0.97 1 0.79 0.61 0.55
mean adv.rad, l2-norm 0.81 0.8 0.78 0.83 0.81 0.79 1 0.63 0.56
mean adv.rad, inf-norm 0.69 0.68 0.66 0.73 0.66 0.61 0.63 1 0.93
mean step rad. inf-norm 0.62 0.62 0.59 0.66 0.6 0.55 0.56 0.93 1
predicted entropy
expected entropy
mutual information
std probas modal class
std probas mean
maxprobs
mean adv.rad, l2-norm
mean adv.rad, inf-norm
mean step rad. inf-norm
1.0
0.8
0.6
0.4
0.2
0.0
Pearson correlation, ensemble trained on CIFAR100
predicted entropy 1 0.99 0.97 0.74 0.99 0.96 0.81 0.68 0.63
expected entropy 0.99 1 0.94 0.7 0.99 0.95 0.8 0.67 0.62
mutual information 0.97 0.94 1 0.81 0.94 0.94 0.81 0.69 0.64
std probas modal class 0.74 0.7 0.81 1 0.73 0.76 0.77 0.79 0.72
std probas mean 0.99 0.99 0.94 0.73 1 0.97 0.84 0.69 0.64
maxprobs 0.96 0.95 0.94 0.76 0.97 1 0.84 0.68 0.63
mean adv.rad, l2-norm 0.81 0.8 0.81 0.77 0.84 0.84 1 0.73 0.76
mean adv.rad, inf-norm 0.68 0.67 0.69 0.79 0.69 0.68 0.73 1 0.95
mean step rad. inf-norm 0.63 0.62 0.64 0.72 0.64 0.63 0.76 0.95 1
predicted entropy
expected entropy
mutual information
std probas modal class
std probas mean
maxprobs
mean adv.rad, l2-norm
mean adv.rad, inf-norm
mean step rad. inf-norm
1.0
0.8
0.6
0.4
0.2
0.0