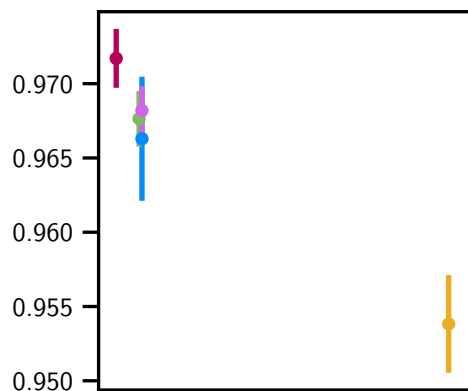
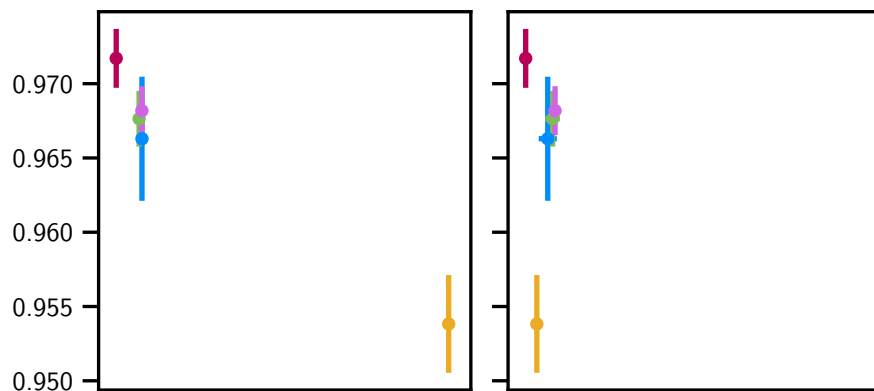
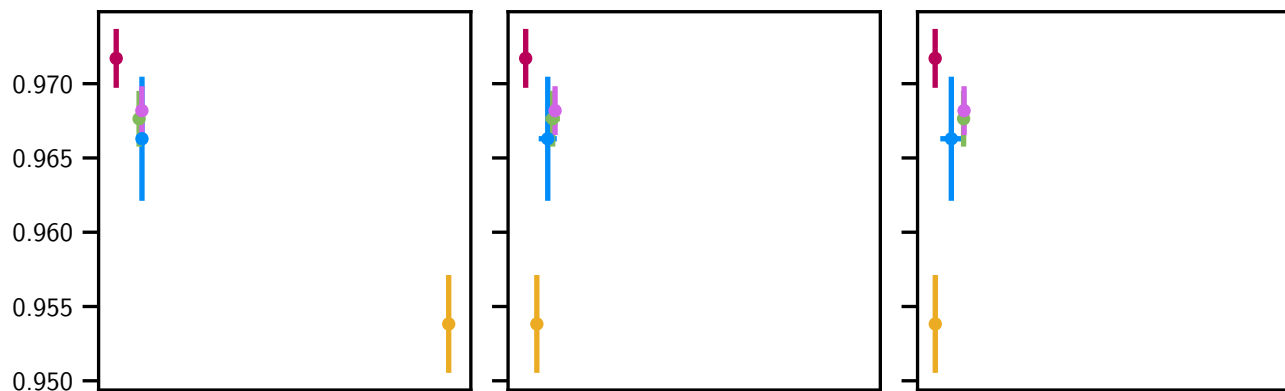
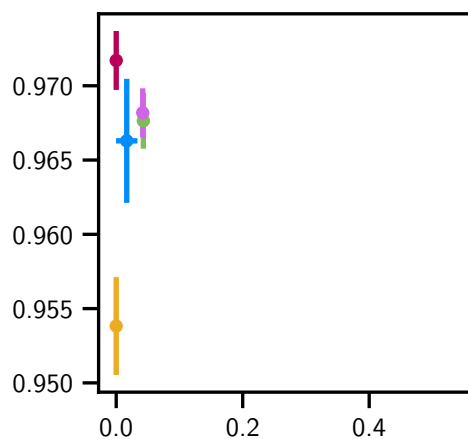
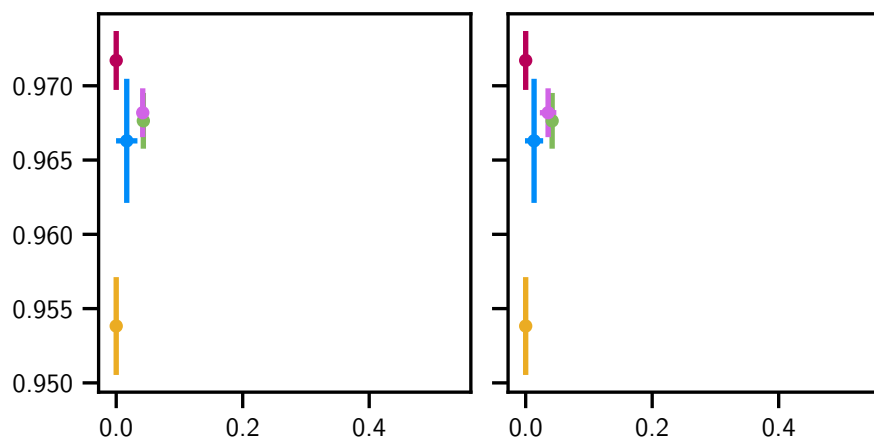


Clean accuracy

PGD evaluation $\epsilon = 0.1$ PGD evaluation $\epsilon = 0.2$ PGD evaluation $\epsilon = 0.3$ PGD evaluation $\epsilon = 0.4$ PGD evaluation $\epsilon = 0.5$ Adversarial training: ϵ

0.0 0.1 0.2
0.3 0.4

Adversarial robustness