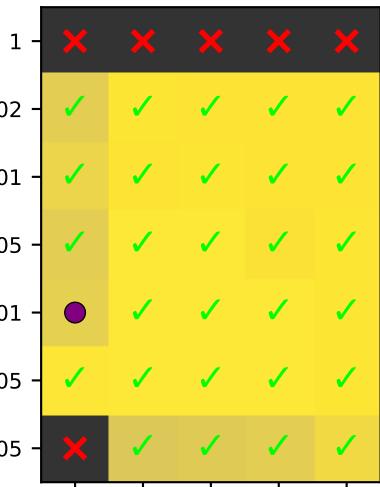
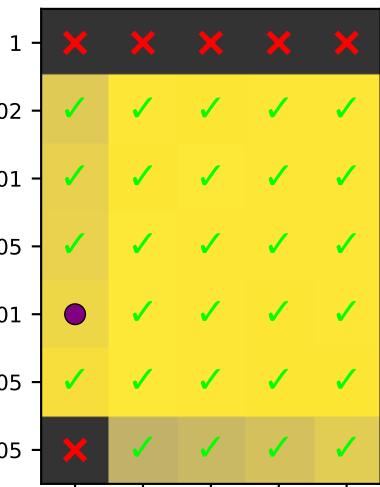
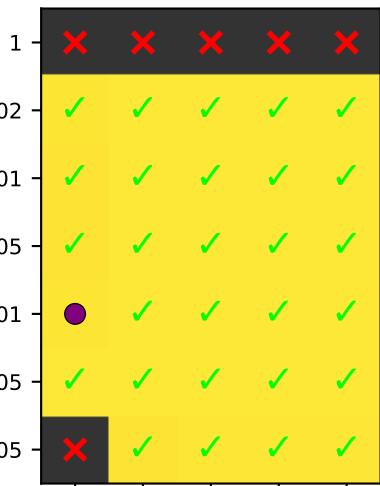


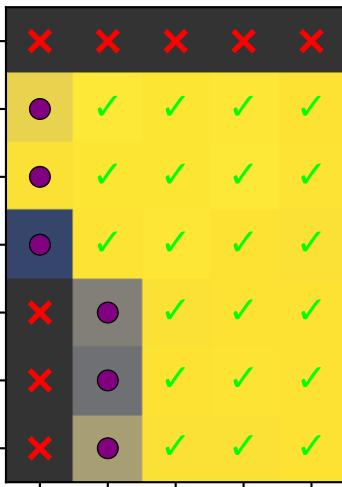
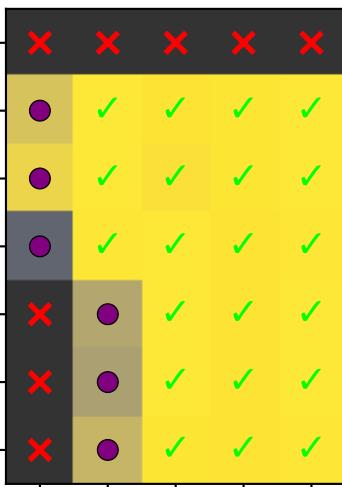
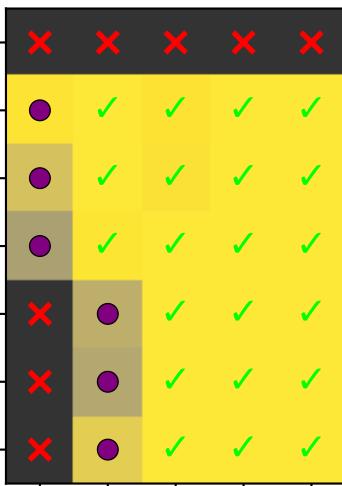
MLP

Transformer

learning rate



weight decay



weight decay