State Tensor Memory Usage Increasing with Model Size Compared to SGD-Boost