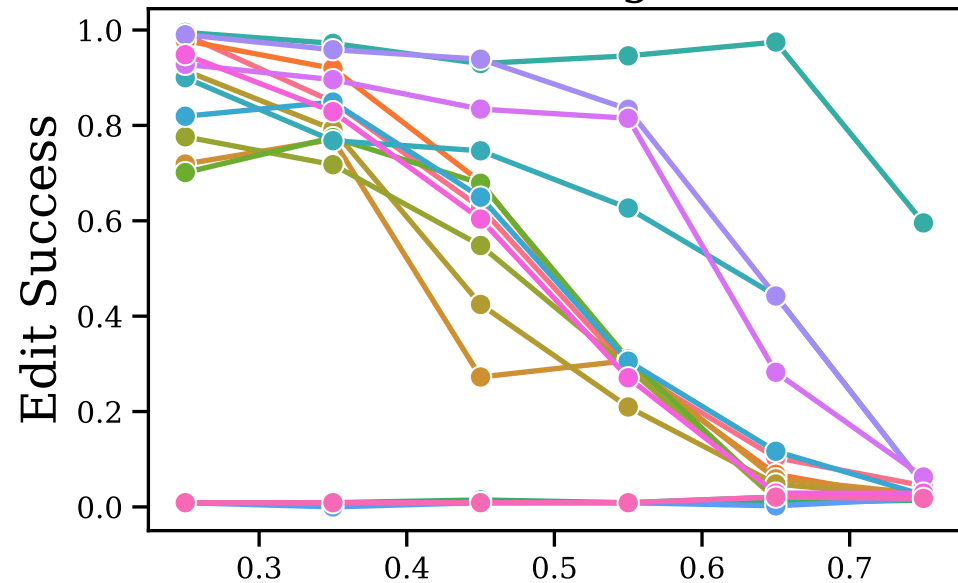
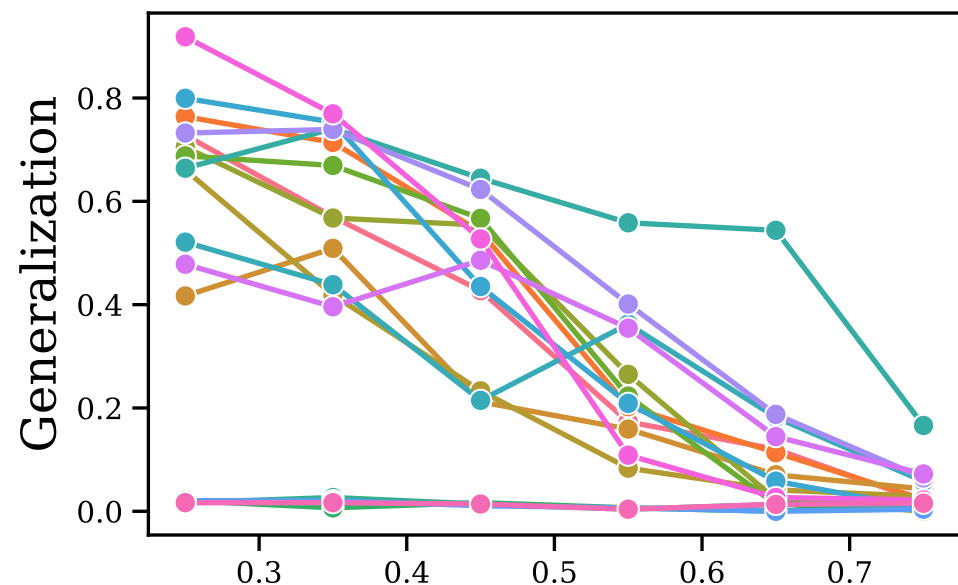


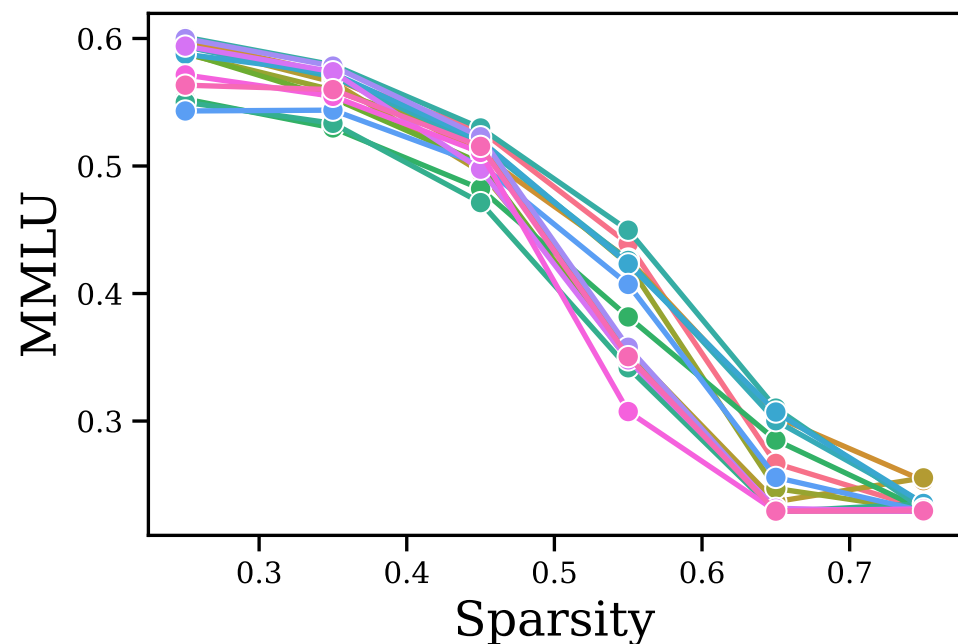
Pruning



Generalization

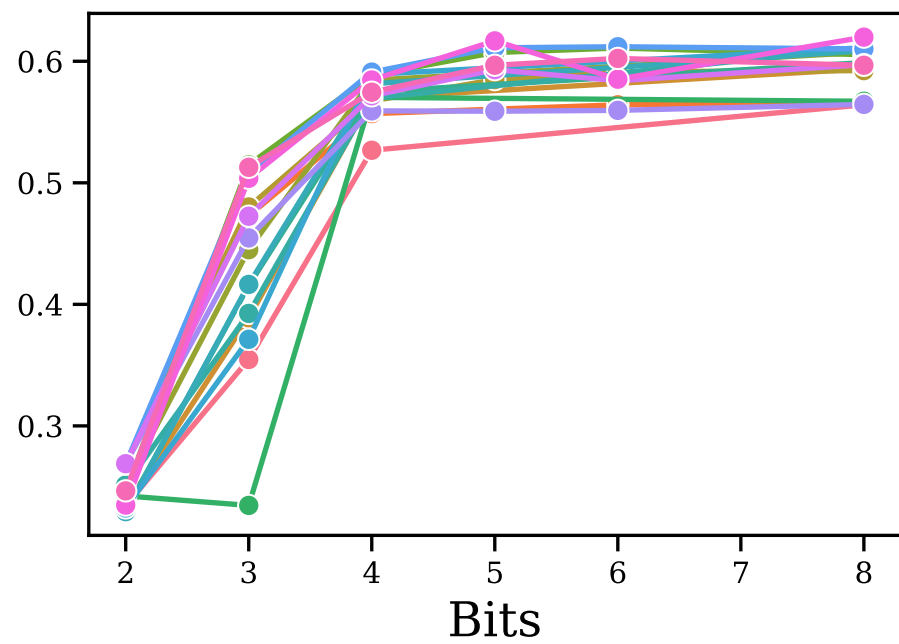
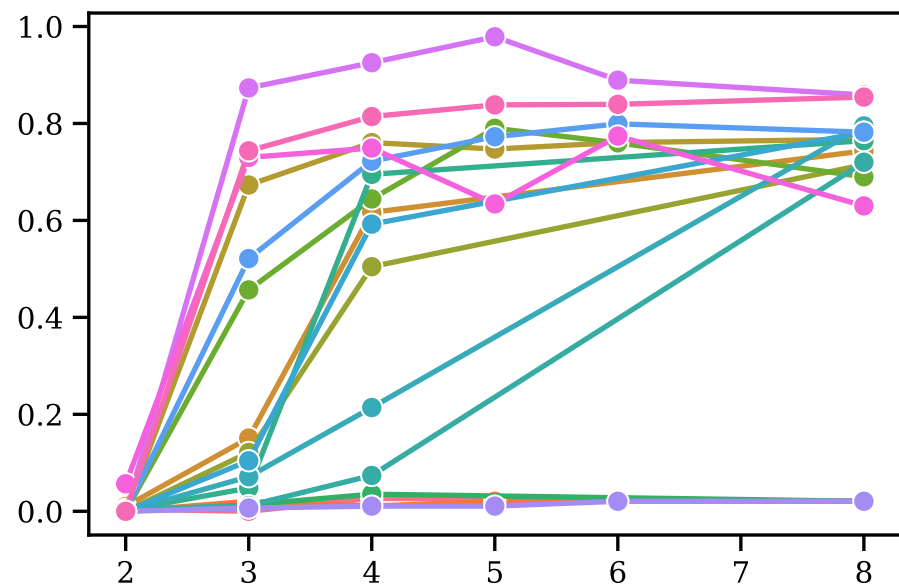
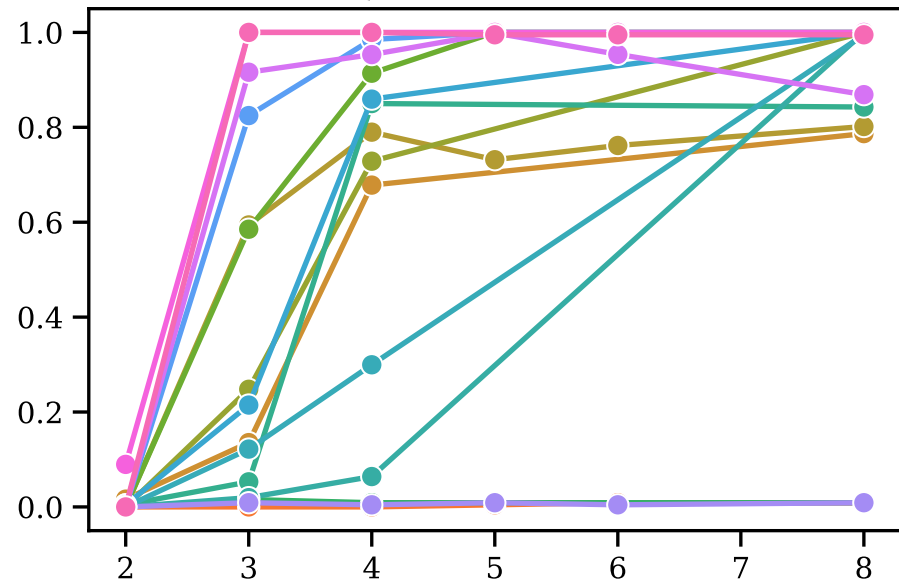


MMLU



Sparsity

Quantization



Bits

