

EEP 596 Quiz 2 – Week 3, Apr 10 (Wed.) Spring 2024

1. How does the number of parameters change when pruning is applied to a neural network, and what is the reason for this change?
2. Which quantization methods does TensorFlow Lite support?
3. Does TensorFlow Lite support structured or unstructured pruning, and what is the key difference between these methods?
4. What distinguishes feature-based knowledge distillation from relation-based knowledge distillation?