## HW 7: ECE 601 Machine Learning for Engineers

**Important Notes:**

(a) If a homework question requires coding, include the full code, program output, and any requested analysis.

(b) Don't panic about the length of the HW assignment. HW assignments are treated as opportunities for improving learning and understanding, so I might include some extra text to help you better understand the concepts or learn about a point that was not covered during the class. The actual work needed from you is indeed manageable.

(c) Submit your solutions as a single zip file named `homework7_UMassUSERNAME.zip`.

In this assignment, you will explore pruning techniques to reduce neural network model size and latency. The objectives are to:

- Understand the fundamentals of **pruning**
- Implement and apply **fine-grained pruning**
- Apply **channel pruning**
- Analyze performance improvements (e.g., speedup) from pruning
- Compare the trade-offs between different pruning approaches

The assignment consists of two main sections: *Fine-Grained Pruning* and *Channel Pruning*, with a total of seven questions:

- Questions 1–5: *Fine-Grained Pruning*
- Question 6: *Channel Pruning*
- Question 7: Comparison of fine-grained and channel pruning

You will use the CIFAR-10 dataset, as in the CNN homework. Follow the provided template, `HW7.ipynb`, which includes step-by-step instructions.