ECE 528: Embedded Systems and Machine Learning

IN

Embedded/IoT systems

- Basic knowledge and hands-on experience with embedded/IoT platforms (e.g., Arduino, Raspberry Pi)

Programming

- Programming experience in Python

Algebra and Calculus

- Basic knowledge of linear algebra and calculus

Algorithms and Data Structures

- Basic knowledge of algorithms and data structures

Processors, Memory, and I/O

- Understand basic structure of a processor, memory subsystem, and I/O organization

Pre-requisites

- ECE251 with a minimum grade of C or CS270 with a minimum grade of C

Concepts:

- Overview of machine learning and embedded systems
- Introduction to deep neural network architectures
  - DNNs, CNNs, GNNs, RNNs, Autoencoders, GANs, …
- Foundations of state-of-the-art deep neural network models
- Model compression techniques for efficient machine learning
  - Sparsification, quantization, knowledge distillation, …
- Hardware acceleration for machine learning on embedded systems
- Hardware/software co-design for machine learning on embedded systems
- Sequence and time series embedded processing
- Foundations of unsupervised learning
- Anomaly detection and security in embedded systems
- Autoencoders and GANs for generative modeling
- Advanced embedded machine learning models and concepts
- Emerging directions and applications in embedded machine learning
- Fairness and bias in machine learning

Applications:

- Deep neural networks for image and video processing
- Deep neural networks for natural language processing
- Deep neural networks for security and anomaly detection
- Deep neural networks for unsupervised learning
- Deep neural networks for generative modeling
- Implementation of deep neural networks on embedded/IoT platforms
- Optimization (software, hardware, codesign) of deep neural networks

Tools:

- Tensorflow, Keras, Jupyter notebooks
- Toolkits for computer vision, natural language processing, …
- Deep learning embedded accelerator hardware simulators
- Embedded platforms/toolkits (e.g., Android smartphones, Google TPU kits)

OUT

- Understand and apply the key design considerations for efficient processing of machine learning models on embedded systems
- Analyze tradeoffs between various embedded software model architectures and optimizations
- Optimize micro-architectural knobs such as precision, data reuse, and parallelism to architect embedded machine learning accelerators given target metrics
- Evaluate the utility of various memory and network design techniques for efficient machine learning processing across embedded system platforms
- Demonstrate skills in hardware/software co-design of machine learning solutions for embedded platforms
- Understand basics of unsupervised learning and its applications to embedded systems
- Analyze challenges with security and adversarial examples for machine learning
- Understand future trends and opportunities from machine learning algorithms, hardware accelerators, and emerging technologies in embedded systems

Revised 1/2023