

שם הקורס: נושאים מתקדמים במבנה מחשבים

מספר הקורס : 236509

סמסטר: אביב תשפ"א

מרצה:	דר' אבי מנדלסון
שעות הרצאה:	
שעת תרגול:	
דרישות קדם:	
אתר הקורס:	

Course name (subtitle): "System Aspects of machine learning"

Target: Efficient implementation of machine learning in general and convolutional networks in particular requires a good understanding of the entire systems; e.g., SW environment to support the development process, hardware capabilities and limitations, optimization techniques and more.

Understanding each of these aspects by itself is not sufficient anymore and an architectural view of the entire system is required in order to educate the student to deal with the challenges of current and future challenges

At the end of the course the student will

1. Have better understanding how different architectures affect the efficiency of CNN algorithm
2. Have experience of using and optimizing CNN applications when running on different systems and development environments
3. Basic knowledge of using CUDA (GPU) and TPU based environments

Grade:

60 – Homework

40 – Final exam / Final project

	Class	Tirgul	Homework	
1	Intro – Machine learning	Python	Setting environment	
2	Intro – CNN, RNN, etc	Pytorch		
3	training	The use of CPUs	Performance and power counters	Measuring CNN training on CPU
4		The use of GPUs	CUDA I	
5		GPU optimizations	CUDA II	Measuring CNN training on GPU
6		Sparsity, pruning and compression	Sparse data structures and algorithms	
7	Inference	Binarization and quantization (SW and algorithms)	Libraries for Binarization and quantization on CPU	Measuring CNN inference on CPU (using FP)
8		The use of CPU, TPU and GPU for low	Libraries for Binarization and	

		precession	quantization on GPUs	
9		Introduction to FPGA	Setting the FPGA environment	Measuring CNN inference on GPU (using quantization)
10		The use of FPGA for binarization and quantization	Running an example on FPGA	
11	In Memory		In Memory Example	Measuring CNN inference on FPGA
12	Deign assistant tools		Example	
13	Put it all together			