



Jae Yun Lee



AI and GPU

홈 > NVDA • NASDAQ

엔비디아

\$891.99

↑108,679.52% +891.17 최대

3월 13일, 오전 10시 6분 30초 UTC-4 · USD · NASDAQ · 면책조항

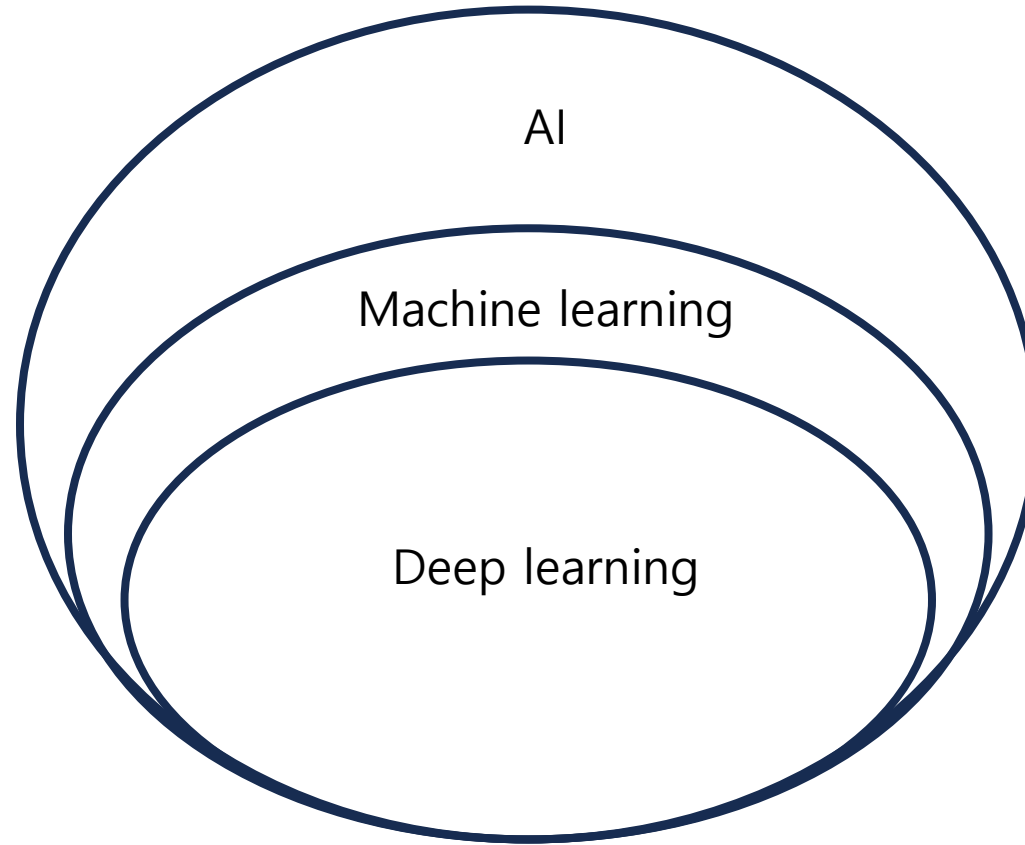
1일 5일 1개월 6개월 YTD 1년 5년 최대



AI and GPU

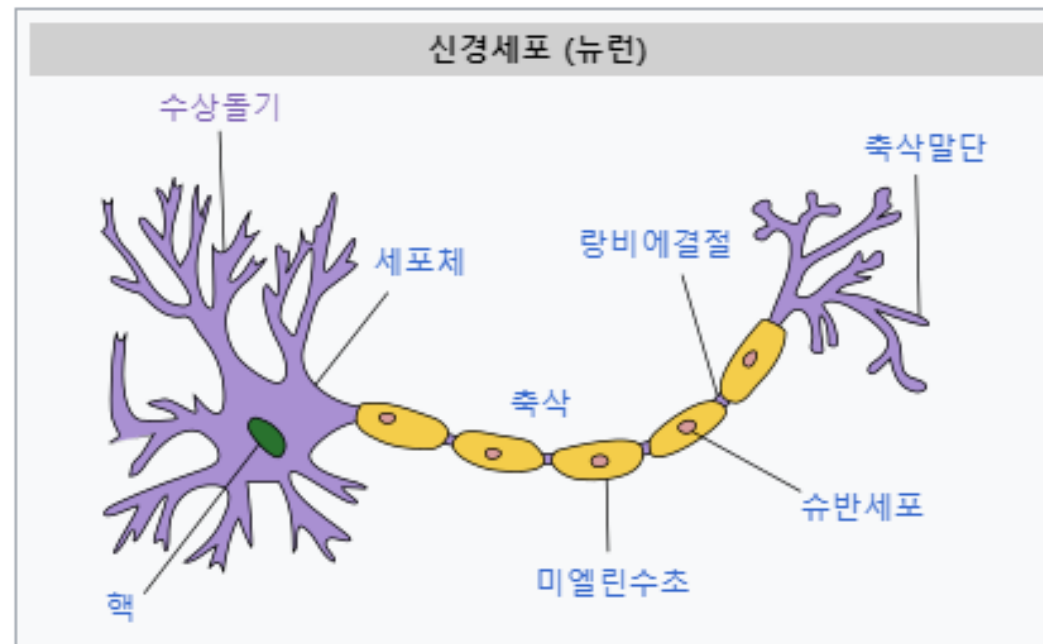
AI 분야에 왜 GPU가 주로 사용될까?

About Deep learning



About Deep learning

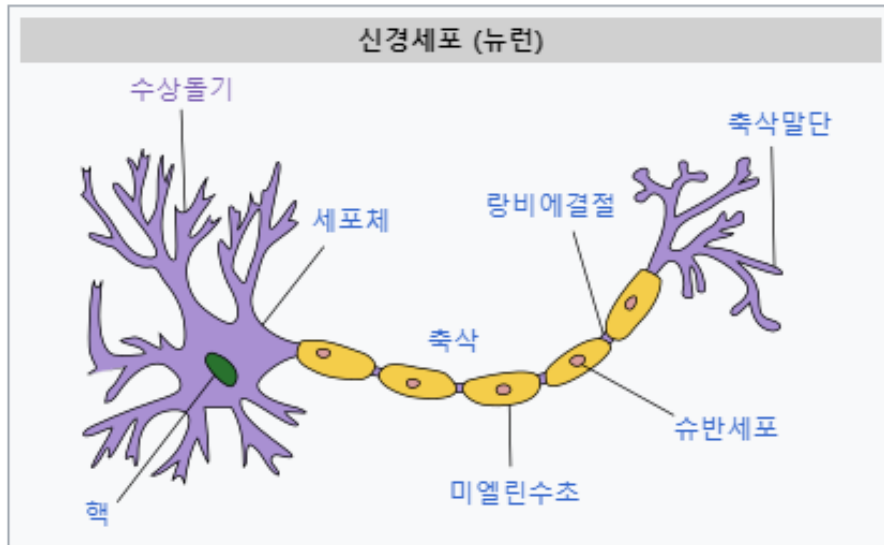
- 신경세포는 **외부의 전달물질**을 세포체에 저장하다가 **자신의 용량을 넘어서면** 축삭돌기를 통해 **외부로 전달물질을 내보냄**.



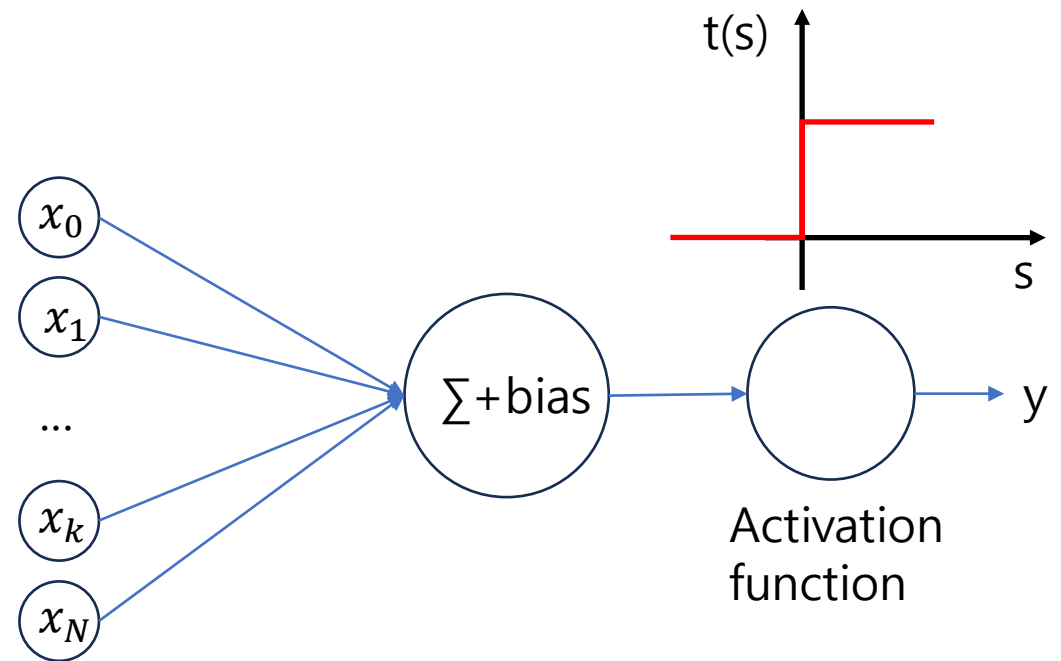
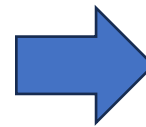
The architecture of Neuron

Single layer Perceptron (SLP)

- Input x 에 대한 weighted sum과 activation function을 통해 output을 도출



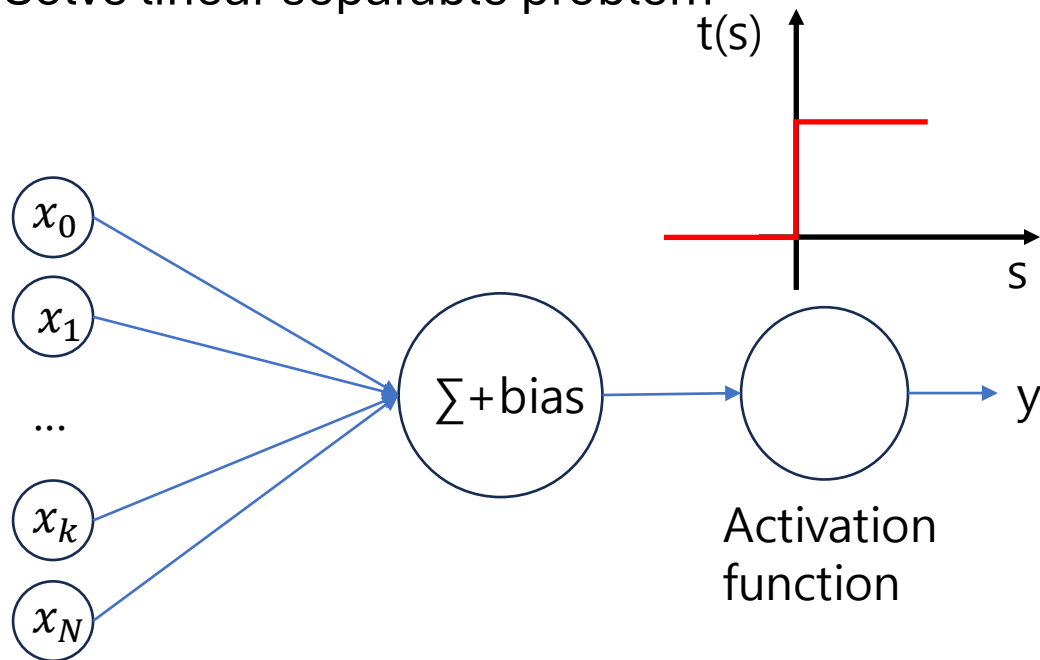
The architecture of Neuron



The example of single layer perceptron

Single layer Perceptron (SLP)

- AND, OR, NAND 문제 해결 가능!
- Solve linear separable problem



The example of single layer perceptron

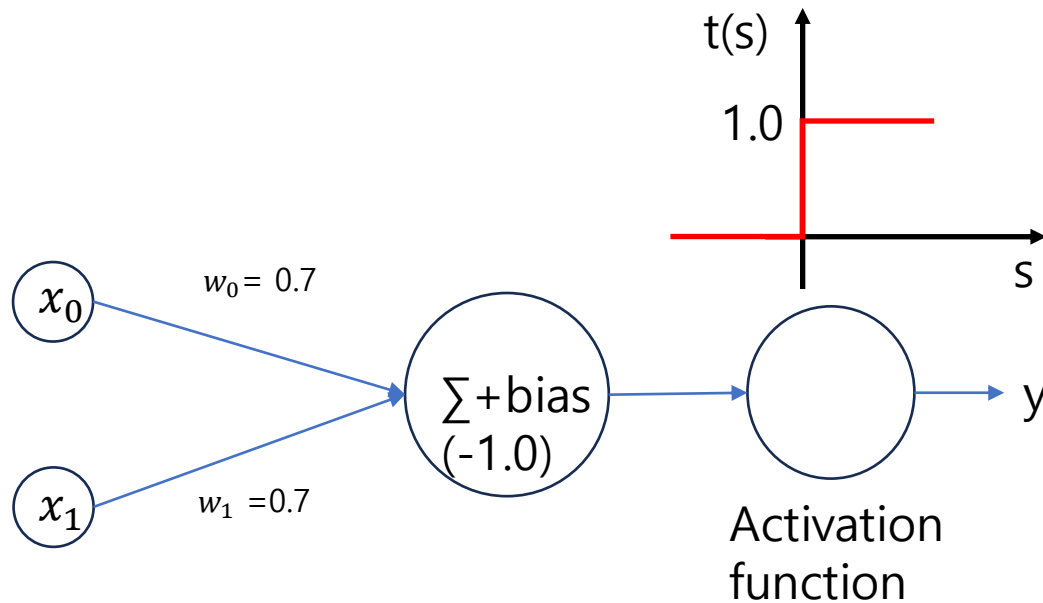
X0	X1	AND	OR	NAND
0	0	0	0	1
0	1	0	1	1
1	0	0	1	1
1	1	1	1	0

The table of AND, OR, NAND gate

Single layer Perceptron (SLP)

- AND problem solution

(*OR과 NAND도 weight와 bias를 조정하여 해결가능)



The example of single layer perceptron

X0	X1	AND	OR	NAND
0	0	0	0	1
0	1	0	1	1
1	0	0	1	1
1	1	1	1	0

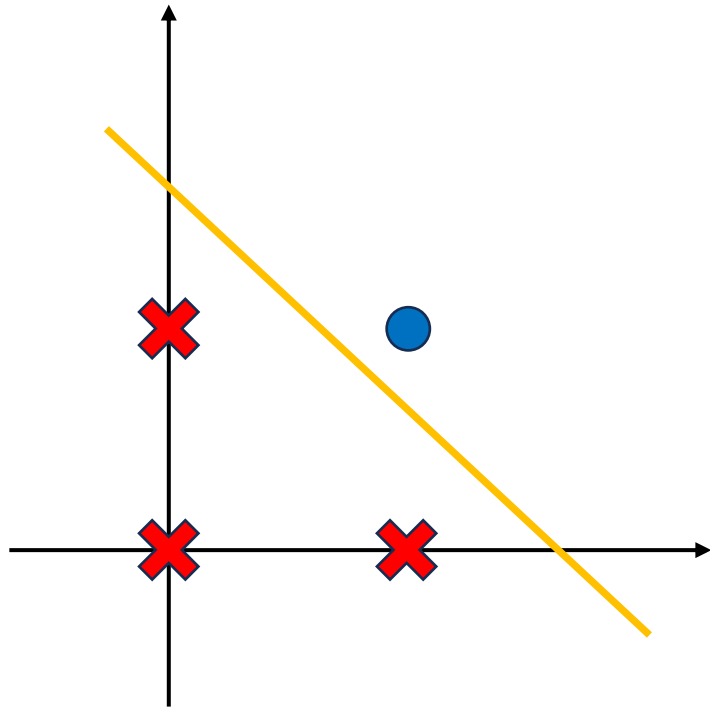
The table of AND, OR, NAND gate

X0	X1	Weighted Sum	Result
0	0	-1.0	0
0	1	-0.3	0
1	0	-0.3	0
1	1	0.4	1

Single layer Perceptron (SLP)

- AND problem solution

(*OR과 NAND도 weight와 bias를 조정하여 해결가능)



The graph of AND gate

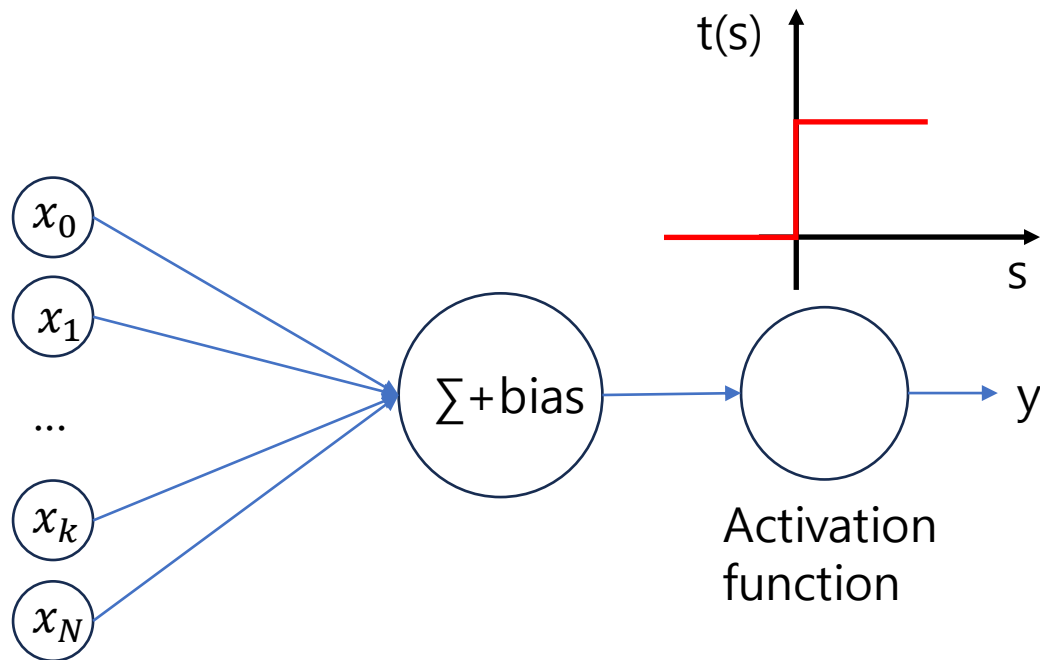
X0	X1	AND	OR	NAND
0	0	0	0	1
0	1	0	1	1
1	0	0	1	1
1	1	1	1	0

The table of AND, OR, NAND gate

X0	X1	Weighted Sum	Result
0	0	-1.0	0
0	1	-0.3	0
1	0	-0.3	0
1	1	0.4	1

Single layer Perceptron (SLP)

- XOR problem solution ?
 - SLP로는 풀 수 없음.

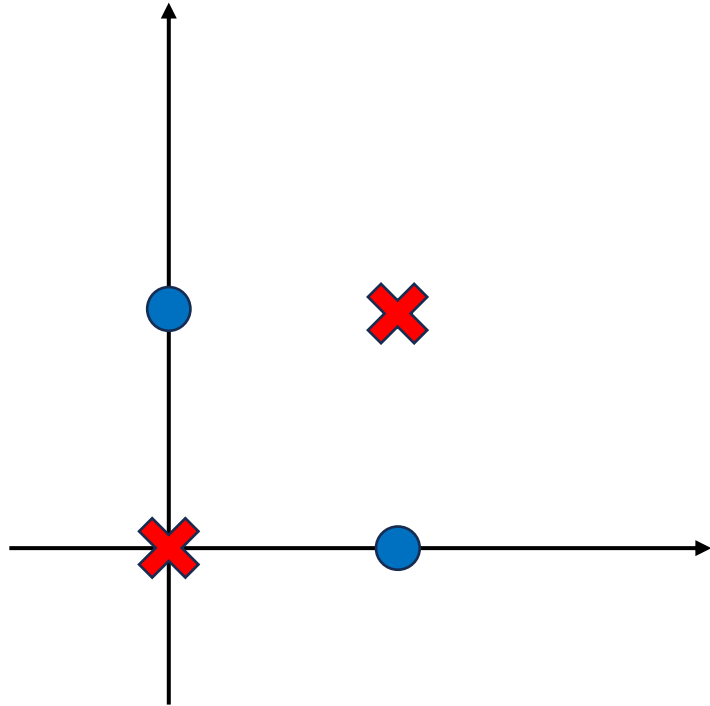


The example of single layer perceptron

X0	X1	XOR
0	0	1
1	0	0
0	1	0
1	1	1

Single layer Perceptron (SLP)

- XOR problem solution ?
 - SLP로는 풀 수 없음.



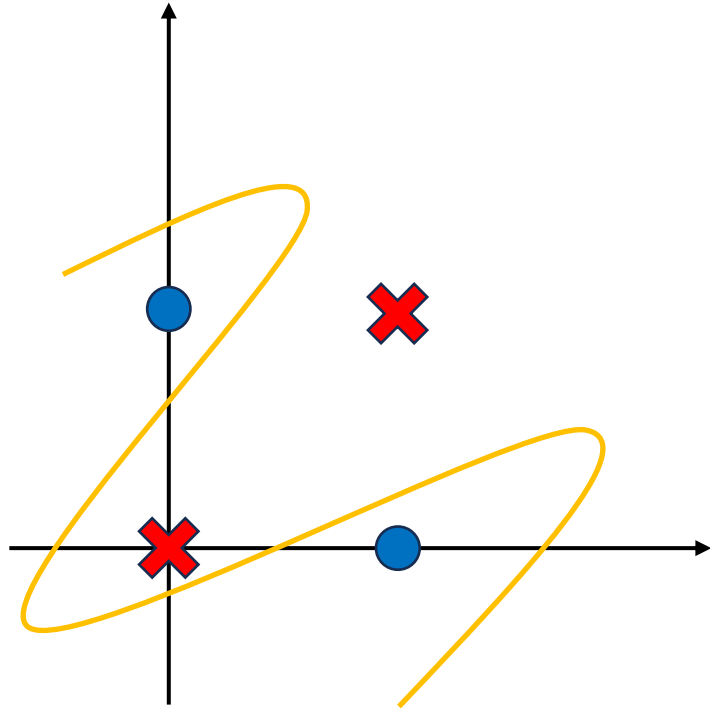
The graph of XOR gate

X0	X1	XOR
0	0	1
1	0	0
0	1	0
1	1	1

The table of XOR gate

Single layer Perceptron (SLP)

- XOR problem solution ?
 - SLP로는 풀 수 없음.



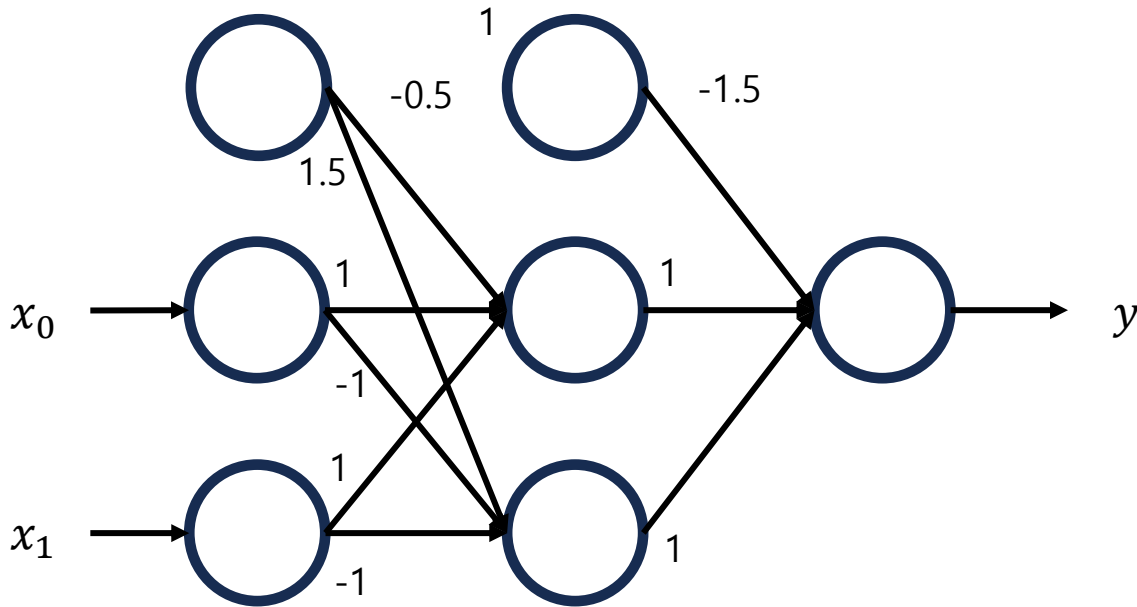
The graph of XOR gate

X0	X1	XOR
0	0	0
1	0	1
0	1	1
1	1	0

The table of XOR gate

Multi layer Perceptron (MLP)

- XOR problem solution
 - SLP를 여러 개 이용



The example of multi-layer perceptron

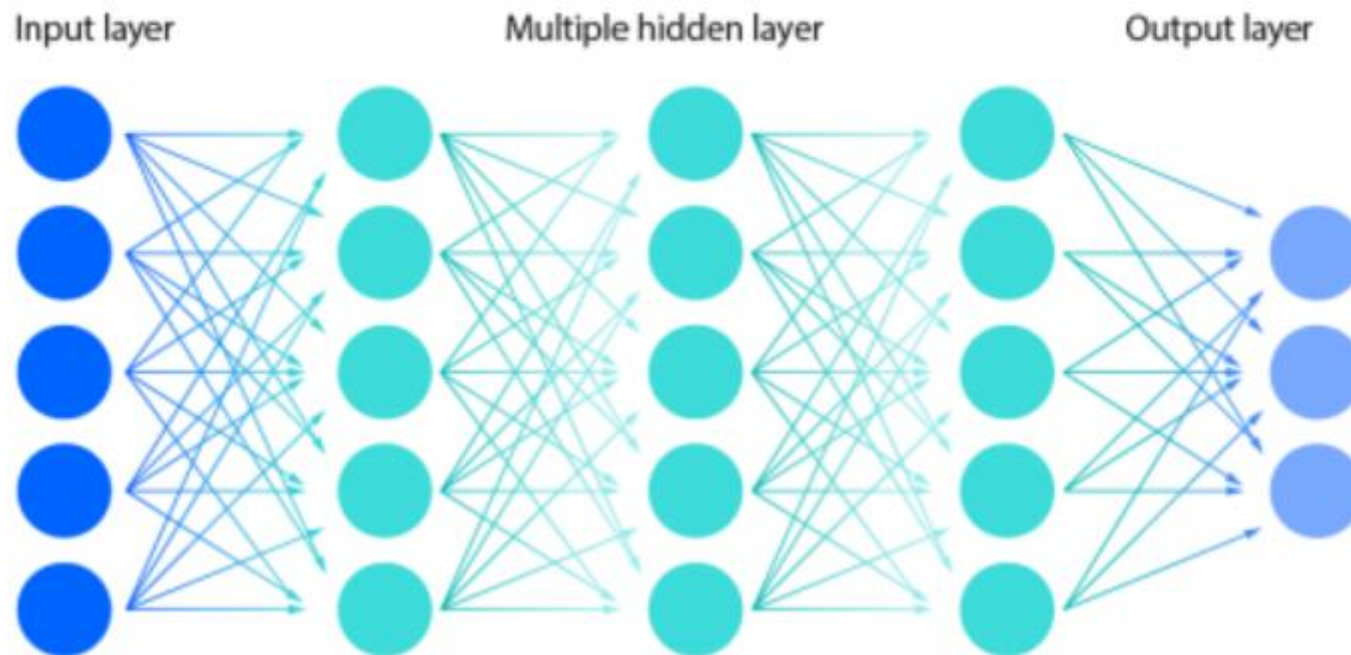
X0	X1	XOR
0	0	0
1	0	1
0	1	1
1	1	0

The table of XOR gate

Deep neural network

- 더 복잡한 Network로 더 복잡한 문제를 풀 수 있다!

Deep neural network



The example image of deep neural network

Why GPU?

- Central processing unit (CPU)

- Control unit

- 기억장치로부터 명령을 꺼내 **순차적으로 해독**
 - 명령어 실행에 필요한 제어신호 전달

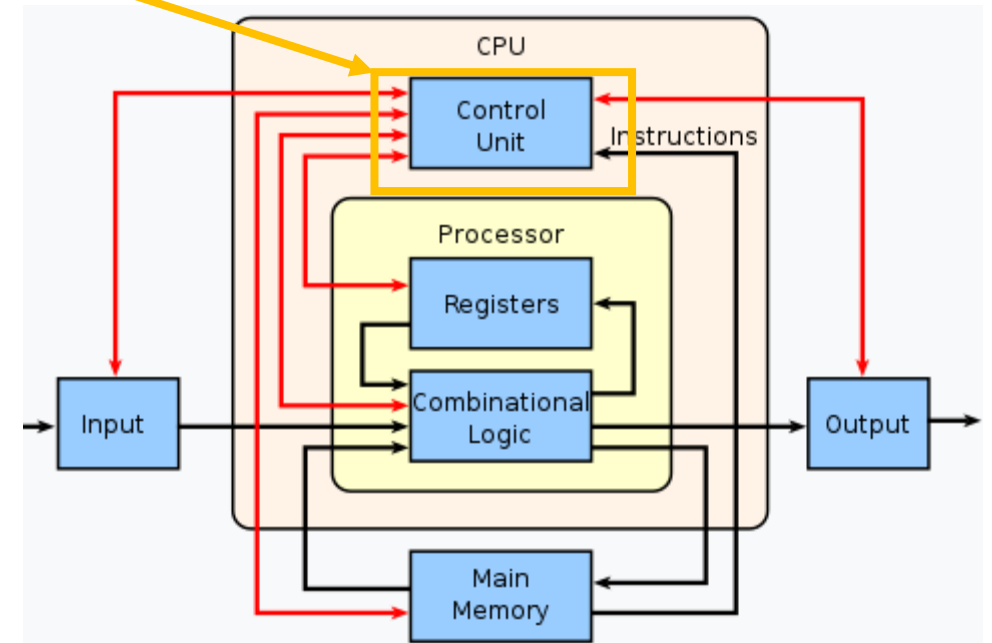
- **한번에 하나의 명령 수행**

- Multi-core CPU

- Core

- 기본 연산과 계산 작업을 수행

- **다수의 작업을 효율적으로 처리**



The architecture of CPU (from wikipedia)

Why GPU?

- Central processing unit (CPU)
 - 3.2GHz = 초당 32억 사이클
(*사이클 : CPU 속도를 측정하는 기본 단위)
 - $24 \times 3.0 \Rightarrow$ 약 720억 사이클

인텔® 코어™ i9-13900K 프로세서 36M 캐시, 최대 5.80GHz	
CPU 사양	
코어 수 ⓘ	24
Performance-core의 수	8
Efficient-core의 수	16
스레드 수 ⓘ	32
최대 터보 주파수 ⓘ	5.80 GHz
Intel® Thermal Velocity Boost 주파수 ⓘ	5.80 GHz
인텔® 터보 부스트 맥스 기술 3.0 주파수 ⓘ	5.70 GHz
Performance-core 최대 터보 주파수 ⓘ	5.40 GHz
Efficient-core 최대 터보 주파수 ⓘ	4.30 GHz
Performance-core 기본 주파수	3.00 GHz
Efficient-core 기본 주파수	2.20 GHz
캐시 ⓘ	36 MB Intel® Smart Cache
총 L2 캐시	32 MB
프로세서 기반 성능 ⓘ	125 W
최대 터보 성능 ⓘ	253 W

The information of intel i9-13900K processor

Why GPU?

- Graphics processing unit (GPU)
 - $10496 \times 1.70 > 17\text{조 } 8430\text{억 사이클}$
- Compare with CPU
 - $720\text{억 사이클 (CPU) vs } 17\text{조 } 8430\text{억 사이클}$

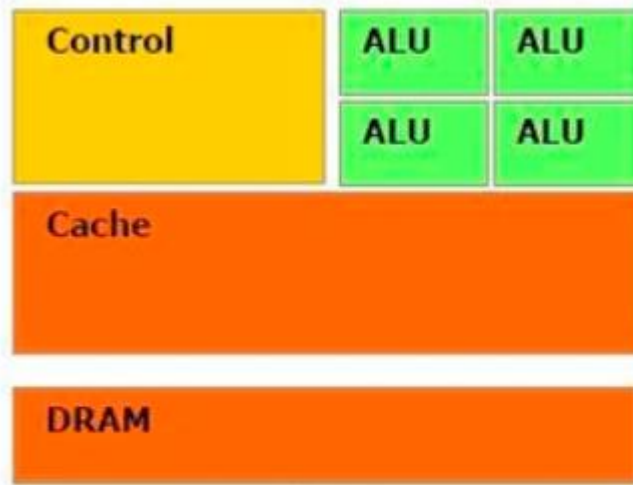


	RTX 3090 Ti	RTX 3090
NVIDIA CUDA 코어	10752	10496
부스트 클럭	1.86 GHz	1.70 GHz
메모리 크기	24 GB	24 GB
메모리 유형	GDDR6X	GDDR6X

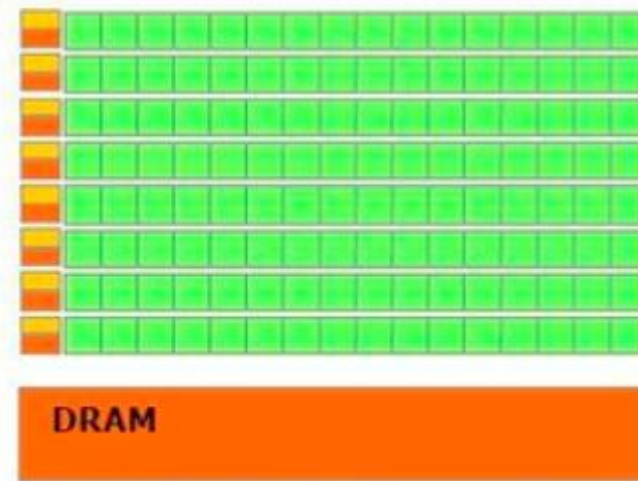
The information of nvidia RTX 3090

Why GPU?

- CPU를 여러 개 늘리면 어떨까?
 - CPU의 코어와 GPU의 코어는 같지 않음.
 - CPU : 연속적인 복잡한 계산
 - GPU : 대량 병렬 연산



CPU

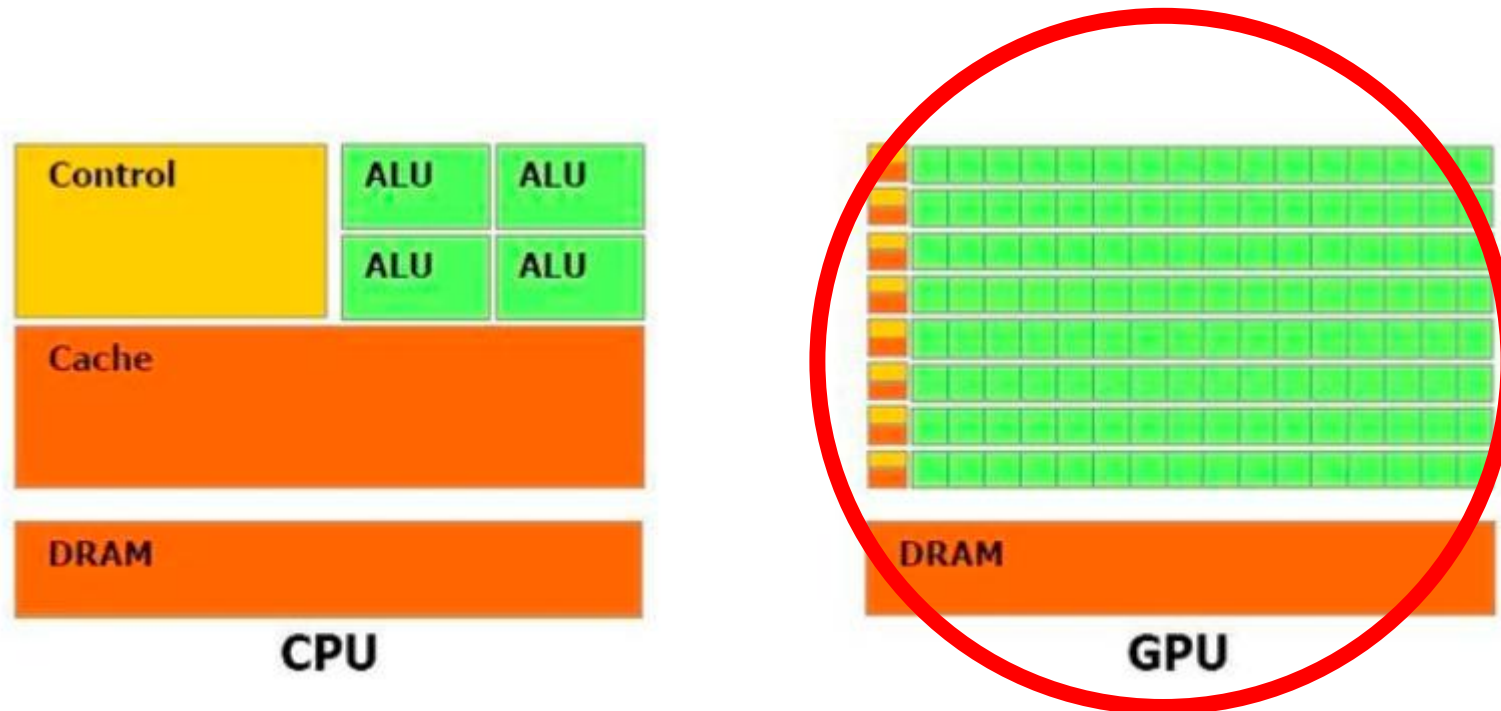


GPU

The architecture of CPU and GPU from SURESOFT

Why GPU?

- Deep neural network를 training 시키고, inference 하기에는 GPU가 적합!



The architecture of CPU and GPU from SURESOFT

AI and GPU

Deep neural network는 기본적으로 많은 weighted sum 연산이 필요
(weighted sum 연산은 다양한 방법으로 수행될 수 있음.)



CPU는 연속적인 복합 계산, GPU는 대량 병렬 연산 가능



단순한 연산인 weighted sum을 병렬적으로 처리하기에는 GPU가 적합!

Thank you

