Machine & Deep learning basics [AICS305]
Machine learning for cyber security II

Prof. Mee Lan Han (aeternus1203@gmail.com)

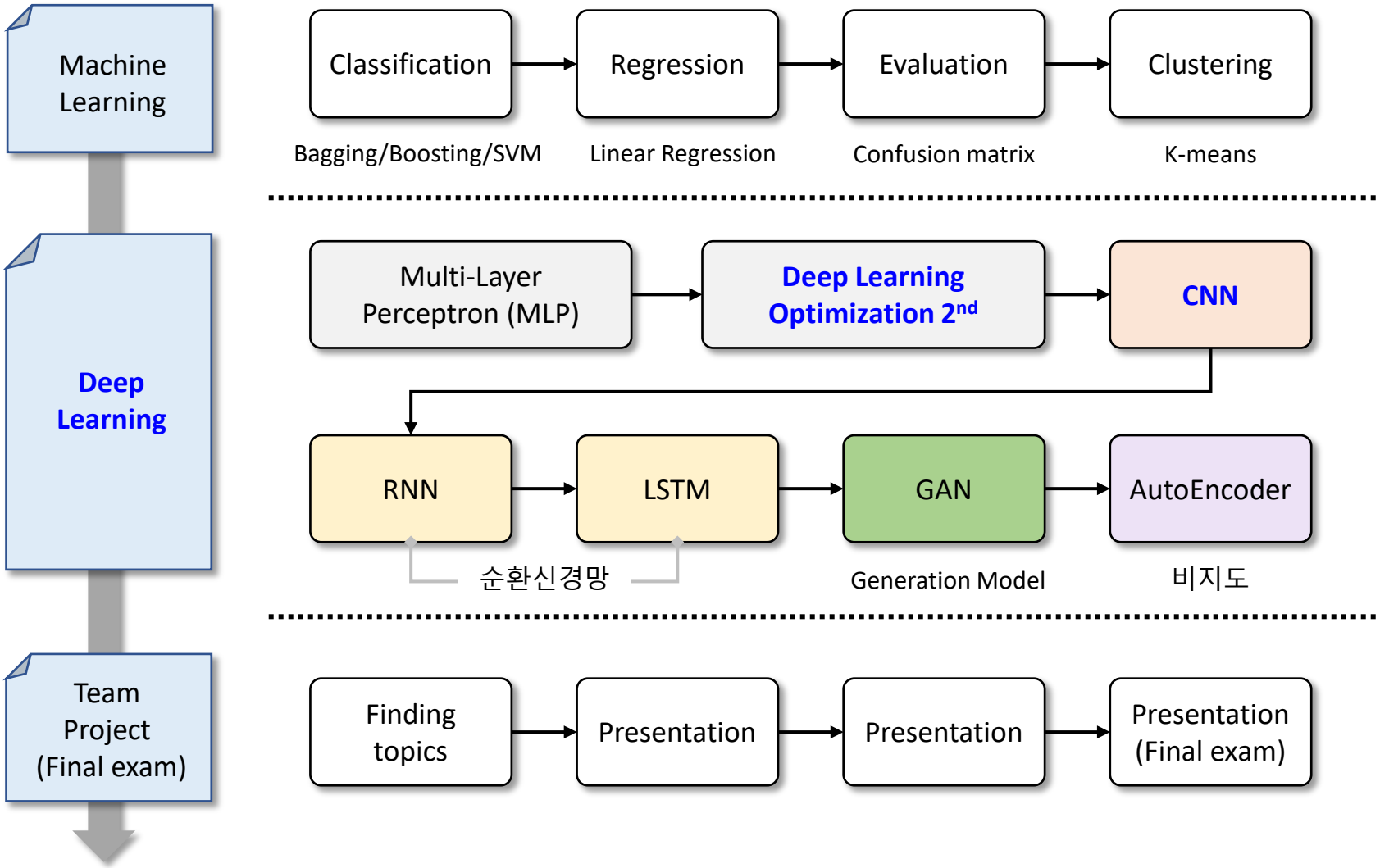
고려대학교

인공지능사이버보안학과

KOREA UNIVERSITY

Machine Learning vs. Deep Learning

■Study Plan



CONTENTS

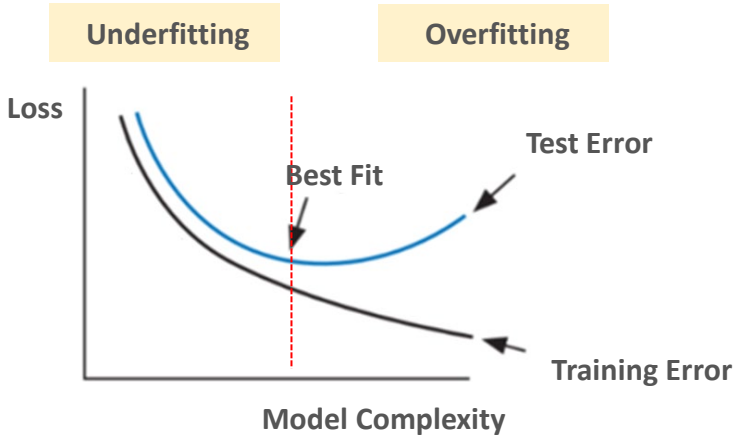
- **Regulation principle**
  - Weight Decay
  - Early Stopping
  - Data Augmentation
  - Dropout

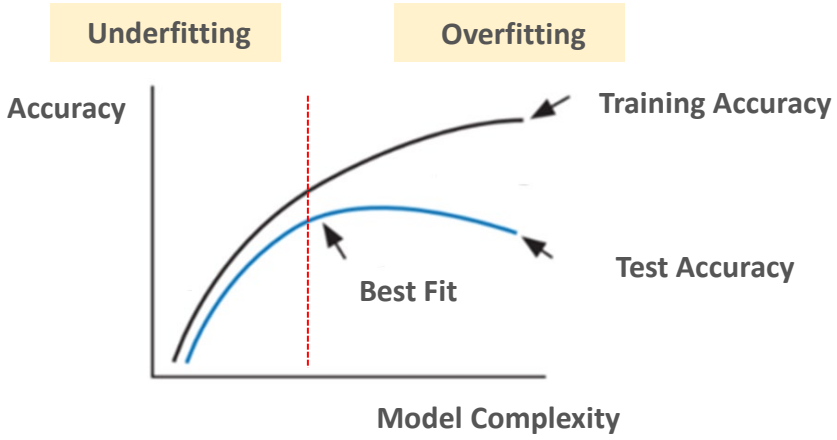
Regulation principle

■ 규제의 필요성

□ 과대적합에 빠지는 이유

- 학습 모델의 용량에 따른 일반화 능력
- Training Dataset을 단순히 '암기' 하는 과대적합에 주의해야 함



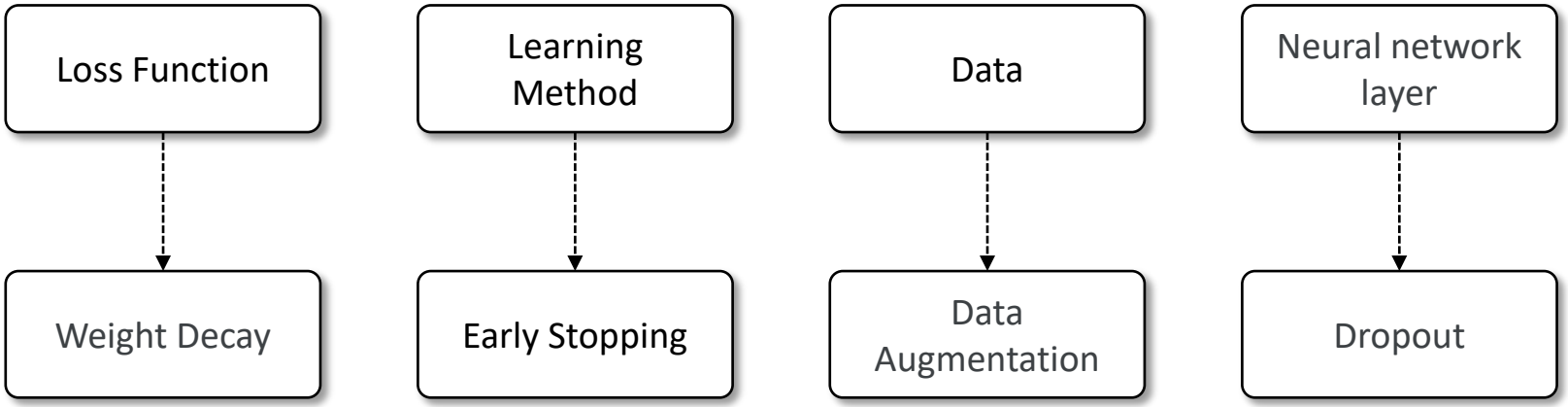


Regulation principle

■ 규제의 필요성

□ 과대적합을 피하는 전략

- 학습 과정에서 여러 **규제 기법**을 적용



Regulation principle

■규제 기법

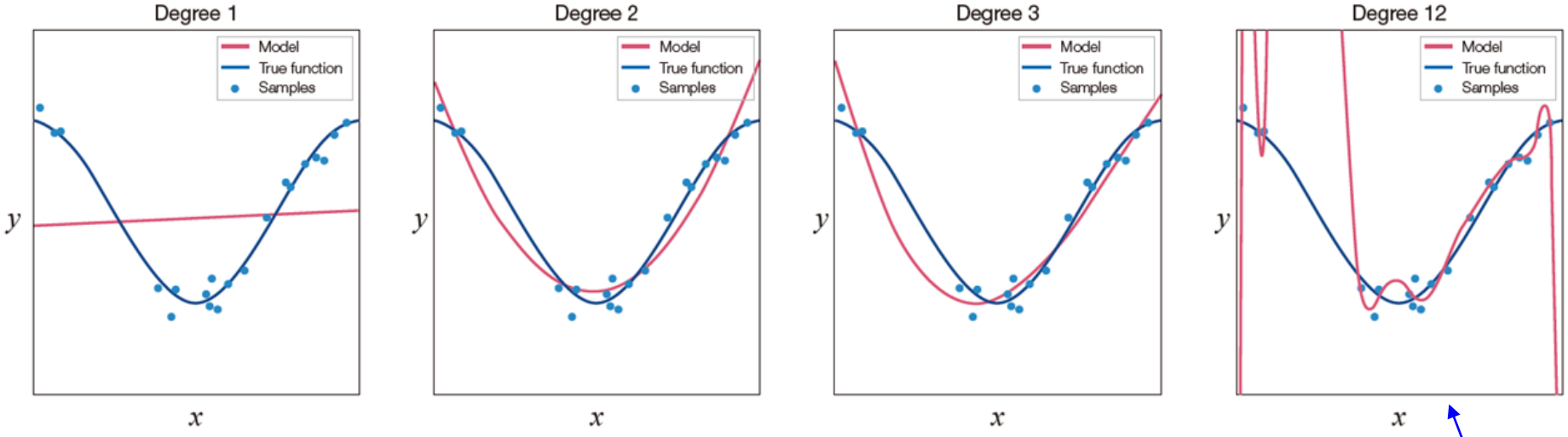
(1) 가중치 감소 (Weight Decay)

□ 과대적합에서는 가중치 값 ($\boldsymbol{\theta}$) 이 아주 큰 현상이 나타남

□ 가중치 감쇠는 성능을 유지한 채로 가중치 크기 ($\boldsymbol{\theta}$)를 낮추는 규제 기법

□ 모델의 weight의 제곱합 (L2 Norm)을 패널티 텀으로 주어 (=제약을 걸어) loss를 최소화 함

= L2 penalty



$y=1005.7x^{12}-27774.4x^{11}+\ldots$

과대적합

Regulation principle

■ 규제 기법

(1) 가중치 감소 (Weight Decay)

- 오차함수에 가중치의 **제곱합 (Norm의 제곱)** 을 더한 뒤, 이를 최소화
- 람다(λ) 는 이 규제화의 강도를 제어하는 파라미터
- 람다를 추가하여 가중치는 자신의 크기에 비례하는 속도로 항상 감쇠하도록 업데이트

$$\underbrace{\boldsymbol{Loss}(\boldsymbol{\theta};\mathbb{X},\mathbb{Y})}_{\text{규제를 적용한 목적함수}} = \underbrace{Loss\ (\boldsymbol{\theta};\mathbb{X},\mathbb{Y})}_{\text{목적함수}} + \underbrace{\frac{1}{2}\gamma\|\boldsymbol{\theta}\|^2}_{\text{규제 항}}$$

- 가중치 $\boldsymbol{\theta}$가 커지게 되면 R항이 커지게 되고 그러면 결과적으로 손실 함수 J가 증가
- 학습 알고리즘은 손실 함수가 작아지도록 학습하므로 R항은 가중치의 크기에 제약을 가하는 역할을 해야 함
- 규제항 R은 가중치를 작은 값으로 유지하므로 모델의 용량을 제한하는 역할
- 규제항은 훈련집합과 무관하며, 데이터 생성 과정에 내재한 사전 지식에 해당
- 규제항 R(theta)로 L2놈이나 L1놈을 사용
  - 큰 가중치에 벌칙, 작은 가중치 유지

Regulation principle

■ 규제 기법

(1) 가중치 감소 (Weight Decay)

□ 규제항 R로 가장 널리 쓰이는 것은 L2 (Norm, 차수) 놈이며 이를 가중치 감쇠 기법이라고 함

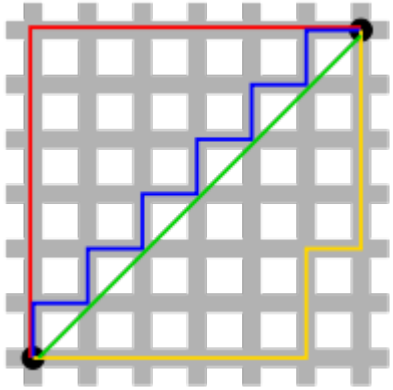
- Norm은 크기의 일반화로 벡터의 크기 (혹은 길이)를 측정하는 방법

$$x = [1, 2, 3, 4, 5]$$
$$||x||_1 = (|1| + |2| + |3| + |4| + |5|)$$
$$= 15$$

L1 Norm은 벡터의 요소에 대한 절댓값의 합

$$x = [1, 2, 3, 4, 5]$$
$$||x||_2 = \sqrt{(|1|^2 + |2|^2 + |3|^2 + |4|^2 + |5|^2)}$$
$$= \sqrt{1 + 4 + 9 + 16 + 25}$$
$$= \sqrt{55}$$
$$= 7.4161$$

L2 Norm은 유클리드 공간에서 벡터 크기 계산



- **L1 Norm**
  - 빨간색, 파란색, 노란색 선으로 표현
  - 여러 가지 path
- **L2 Norm**
  - 오직 초록색 선으로만 표현
  - Unique shortest path

Regulation principle

■ 규제 기법

(1) 가중치 감소 (Weight Decay)

□ 규제항 R로 가장 널리 쓰이는 것은 L2 (Norm, 차수) 놈이며 이를 가중치 감쇠 기법이라고 함

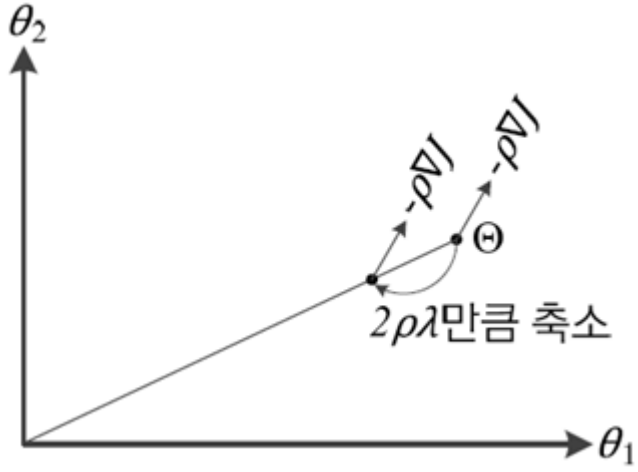
□ 매개변수 갱신하는 수식

$$\boldsymbol{Loss}(\boldsymbol{\theta}; \mathbb{X}, \mathbb{Y}) = Loss\ (\boldsymbol{\theta}; \mathbb{X}, \mathbb{Y}) + \frac{1}{2}\gamma \|\boldsymbol{\theta}\|^2$$

Loss에 대한 미분

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta}\ - \eta\ (\frac{\partial DataLoss}{\partial \boldsymbol{\theta}} + \lambda \boldsymbol{\theta})$$

$$= \boldsymbol{\theta}(\mathbf{1}\ - \eta\lambda) - \eta(\frac{\partial DataLoss}{\partial \boldsymbol{\theta}})$$



최종 해를 원점 가까이 당기는 효과
(즉, 가중치를 작게 유지함)

□ 미분을 했을 때, 기본 Dataloss 에 $\boldsymbol{\theta}$ 의 lambda배 만큼을 더하게 되므로 가중치 값이 그만큼 보정

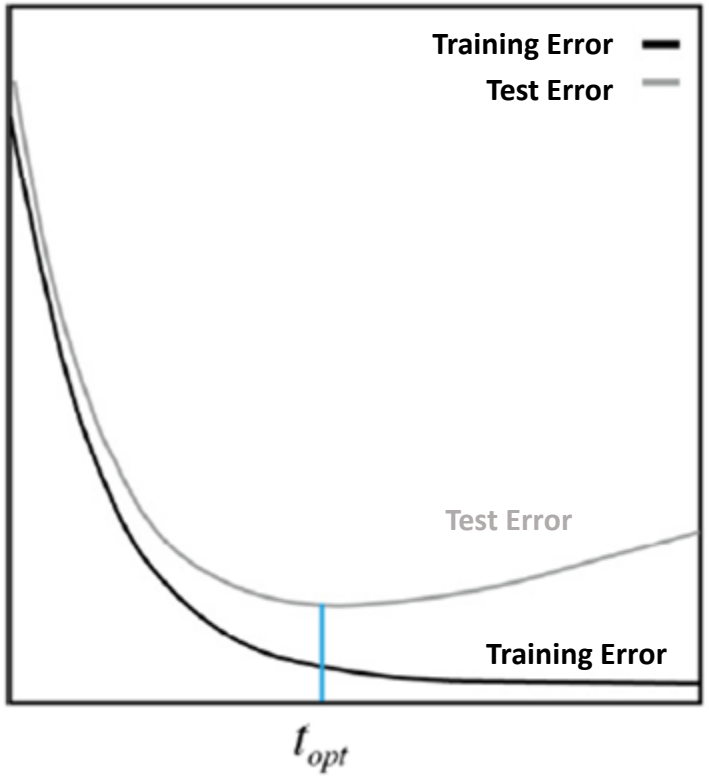
□ $\boldsymbol{\theta}$(1-학습률*lambda)가 되기 때문에 weight가 아주 작은 factor에 비례하여 감소함

Regulation principle

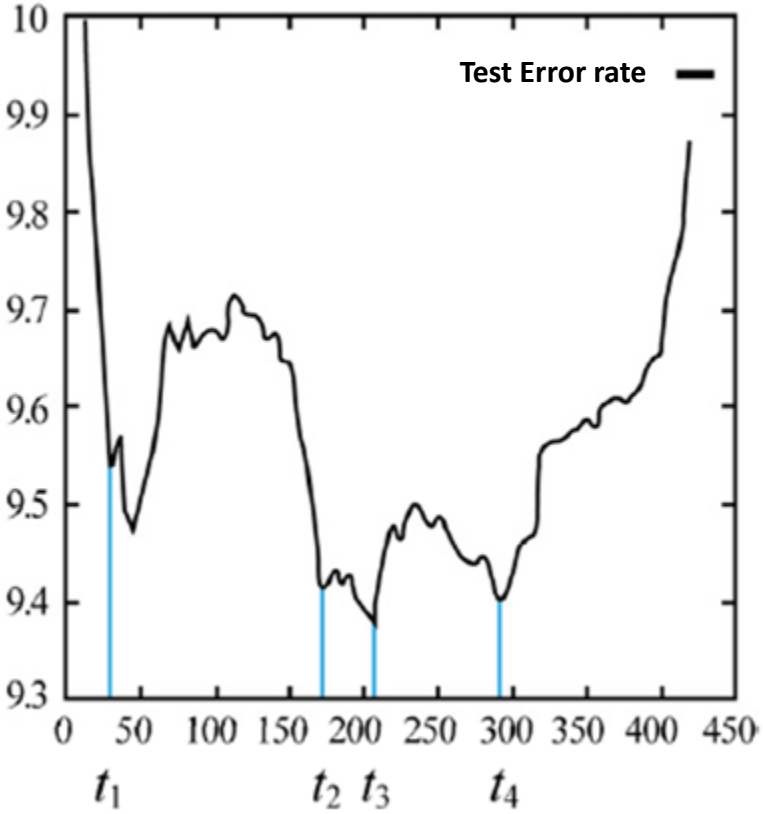
■ 규제 기법

(2) 조기 종료 (Early Stopping)

- □ 일정 시간(t_{opt})이 지나면 과대적합 현상이 나타남 → 일반화 능력 저하
- □ 즉 훈련 데이터를 단순히 암기하기 시작



검증집합의 오류가 최저인 점 t_{opt} 에서 학습을 멈춤



실제 데이터에 나타나는 지그재그 현상

Regulation principle

■ 규제 기법

(2) 조기 종료 (Early Stopping)

□ 모델이 과적합되기 전 훈련을 멈추는 정규화 기법
□ 훈련 중 주기적으로 성능검증 수행, 성능이 더 좋아지지 않으면 과적합이라 판단하고 훈련 멈춤
□ Epoch 단위로 성능 검증 수행, Epoch보다 자주 검증해야 할 때는 Batch 실행 단위로 검증하기도 함

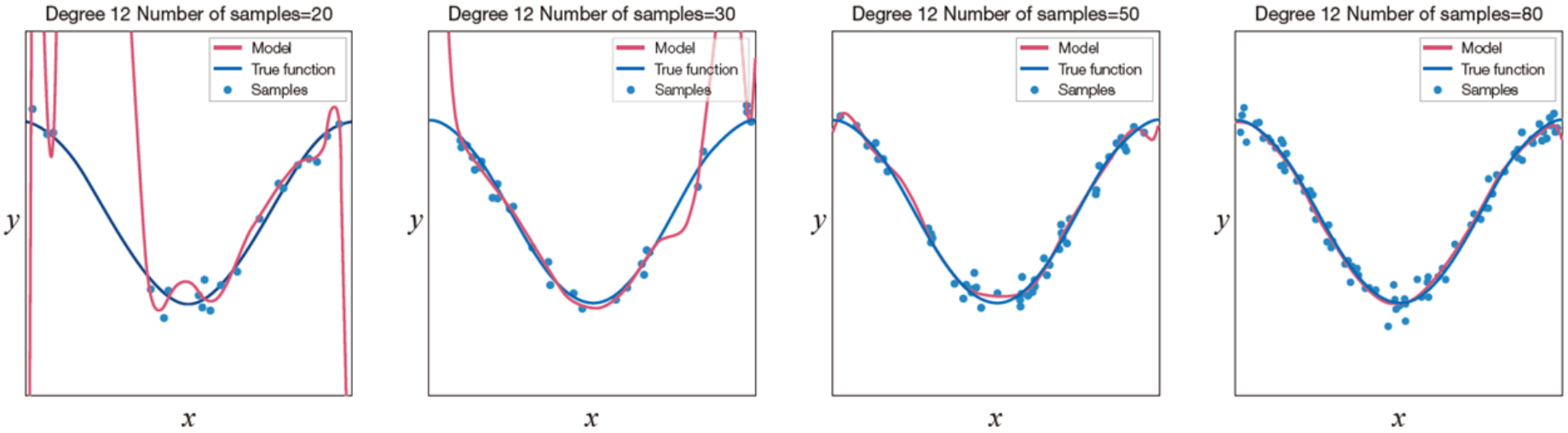
□ 조기 종료 기준
- 모델의 성능이 바로 향상하지 않는다고 종료해버리면 학습이 제대로 되지 않을 수 있음
- 일시적 변동이 아닌 지속적인 정체 또는 하락에 의한 판단이 들었을 때 종료!!

Regulation principle

■규제 기법

(3) 데이터 증대 (Data Augmentation)

- □ 과대적합을 방지하는 가장 확실한 방법은 큰 훈련 집합 사용
- □ 대부분 상황에서 데이터를 늘리는 일은 많은 비용이 소요
- □ 딥러닝에서는 주어진 데이터를 인위적으로 늘리는 데이터 증대/증강 (data augmentation)를 적용
  - 영상을 이동, 회전 또는 좌우 반전
  - 명암 조정 등
  - 텐서플로는 훌륭한 함수 제공하고 있음



Size of a training dataset: 20 | Size of a training dataset: 30 | Size of a training dataset: 50 | Size of a training dataset: 80

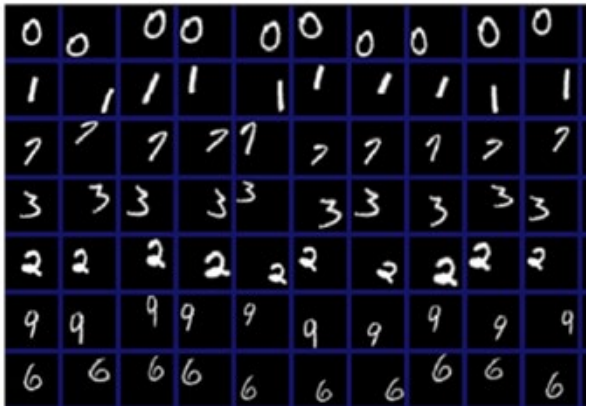
데이터가 커지면 과대적합이 자연스럽게 사라지는 현상 (예시)

Regulation principle

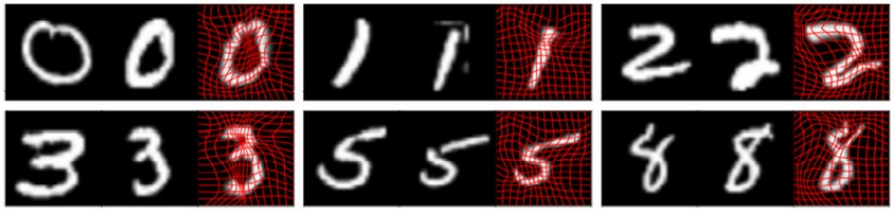
■규제 기법

(3) 데이터 증대 (Data Augmentation)

- MNIST에 Affine 변환 (이동(Translation), 회전(Rotation), 확대(Zoom), 반전(Invert), 전단(Shearing))을 적용



- 모핑 (Morphing)을 이용한 변형
  - ✓ 비선형 변환으로서 어파인 변환에 비해 훨씬 다양한 형태의 확대
  - ✓ 학습 기반: 데이터에 맞는 비선형 변환 규칙을 학습

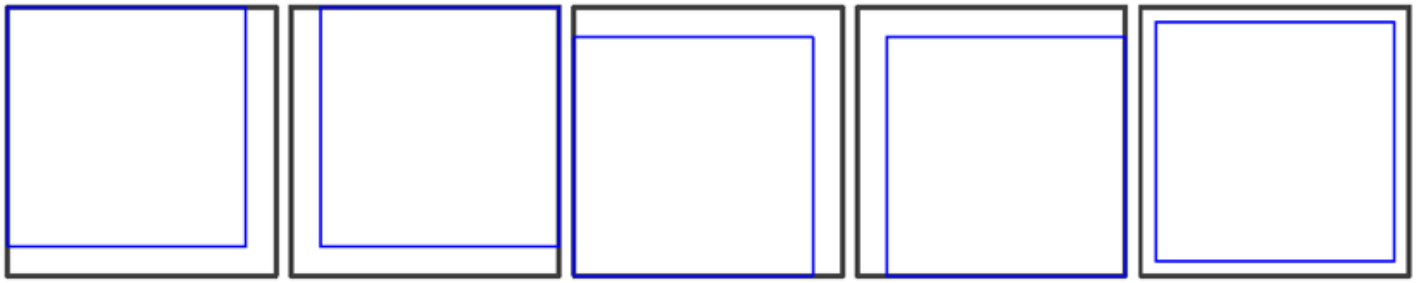


Regulation principle

■ 규제 기법

(3) 데이터 증대 (Data Augmentation)

- 자연영상 확대
  - ✓ 256*256 영상에서 224*224 영상을 1024장 잘라내어 이동 효과. 좌우 반전까지 시도하여 2048배로 확대
  - ✓ PCA를 이용한 색상 변환으로 추가 확대
  - ✓ 예측 단계에서 5장 잘라내고 좌우 반전하여 10장을 만든 다음 앙상블 적용



예측 단계에서 영상 잘라내기

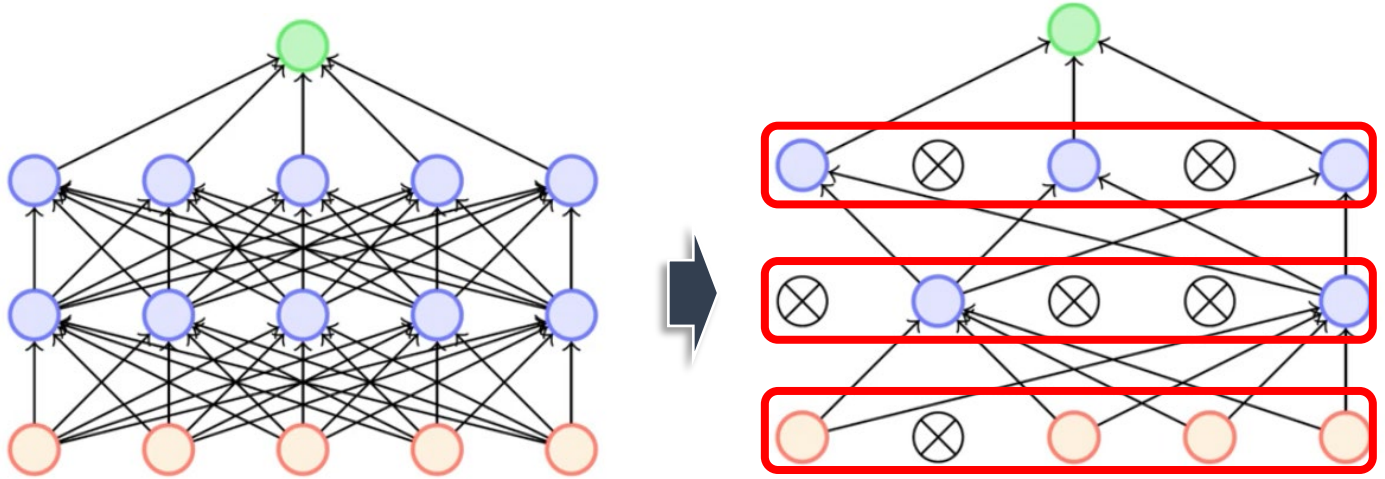
- 잡음을 섞어 확대하는 기법
  - ✓ 입력 데이터에 잡음을 섞는 기법
  - ✓ 은닉 노드에 잡음을 섞는 기법 (고급 특징 수준에서 데이터를 확대하는 셈)

Regulation principle

■ 규제 기법

(4) Dropout

□ 신경망 전체를 다 학습시키지 않고 **일부 노드만 무작위로 골라 학습**시키는 기법
□ 일정 비율의 가중치를 임의로 선택하여 불능으로 만들고 학습하는 규제 기법
□ 학습하는 중간에 일정 비율로 노드들의 출력을 0으로 만들어 신경망의 출력을 계산함
(**특정 뉴런의 확률 p를 0으로 바꾸는 것을 의미**)



□ Dropout 적용 순서: ReLU등의 Activation 함수 적용 후, Pooling 이전일때가 가장 적절
- Convolution -> Batch Normalization -> Activation -> **Dropout** -> Pooling

Thank you

KOREA
UNIVERSITY