

Part.04

Ensemble Learning

| DS분야에서 Tree기반 모델이 쓰이는 이유

FASTCAMPUS
ONLINE

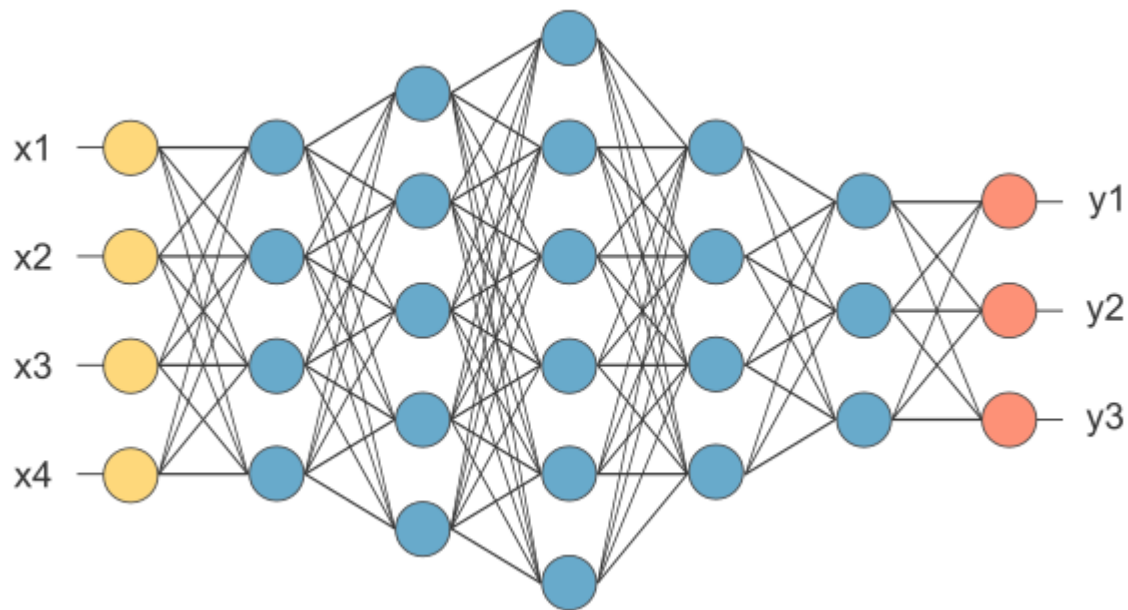
머신러닝과 데이터분석 A-Z

강사. 이경택

IDS분야에서 Tree기반 모델이 쓰이는 이유

■ Neural Network

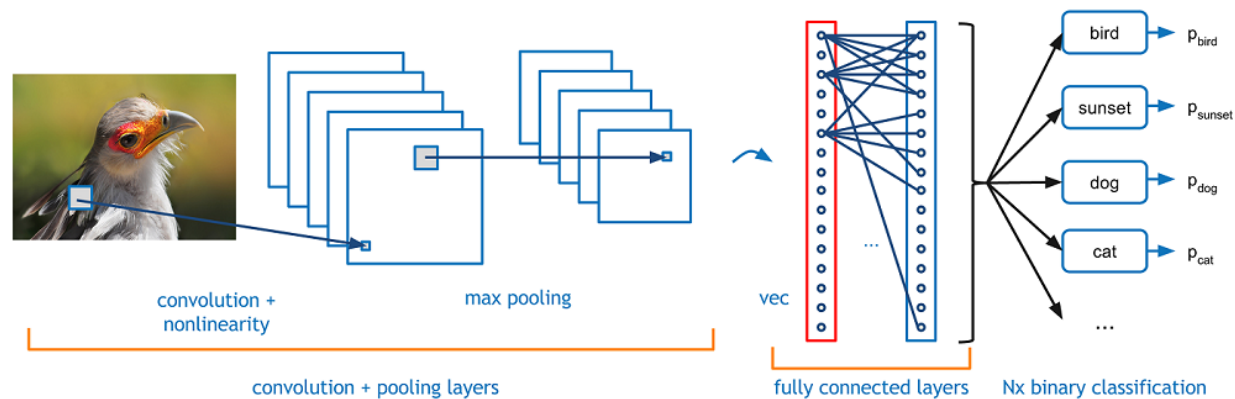
- 입력, 은닉, 출력층으로 구성된 모형으로서 각 층을 연결하는 노드의 가중치를 업데이트하면서 학습
- Overfitting이 심하게 일어나고 학습시간이 매우 오래걸림



I DS분야에서 Tree기반 모델이 쓰이는 이유

■ Deep Learning

- 다층의 layer를 통해 복잡한 데이터의 학습이 가능토록 함 (graphical representation learning)
- 알고리즘 및 GPU의 발전이 deep learning의 부흥을 이끔



- 이미지 처리에 사용되는 CNN(Convolutional Neural Network)

IDS분야에서 Tree기반 모델이 쓰이는 이유

- 이미지 분류에서 기존 모델
 - 각각의 픽셀 값(늘어뜨려서)을 독립변수로 사용
 - 독립변수들은 각각 독립이라는 기본적인 가정에서 어긋남

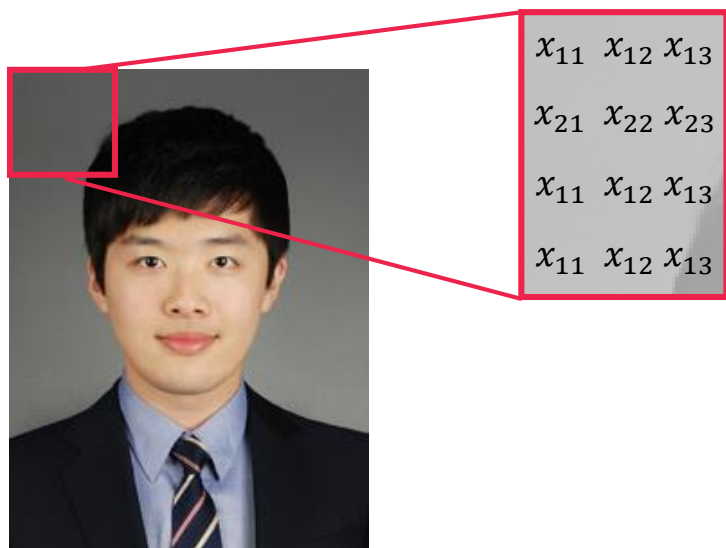


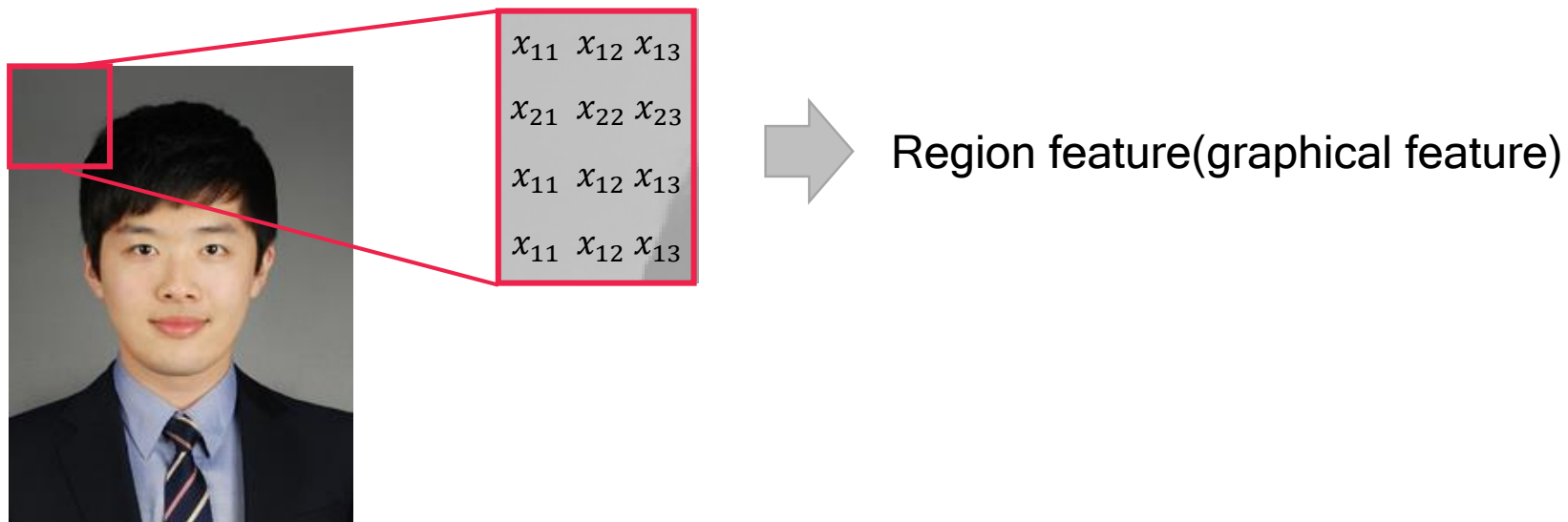
Image1
Image2
...

V1	V2	V3	V4	V5	...	V11	V12	V13	Y
x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	...	x_{11}	x_{12}	x_{13}	사람
									고양이

기존 Machine Learning

IDS분야에서 Tree기반 모델이 쓰이는 이유

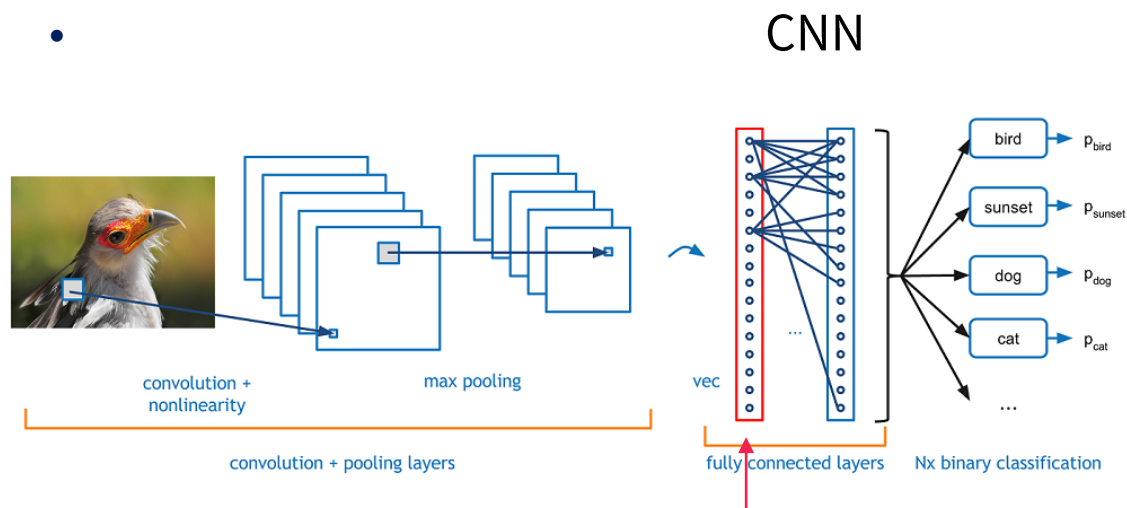
- Convolutional Neural Network
 - 이미지의 지역별 feature를 뽑아서 neural network학습



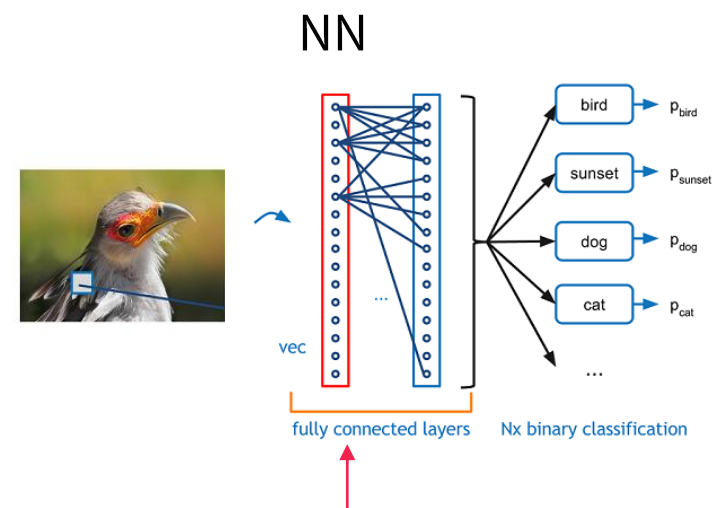
I DS분야에서 Tree기반 모델이 쓰이는 이유

Convolutional Neural Network

- 이미지의 지역별 feature를 뽑아서 neural network학습
- CNN의 해주는 역할은 이미지의 feature를 잘 뽑기 위한 것 뿐
-



Convolution을 통해 뽑은 feature



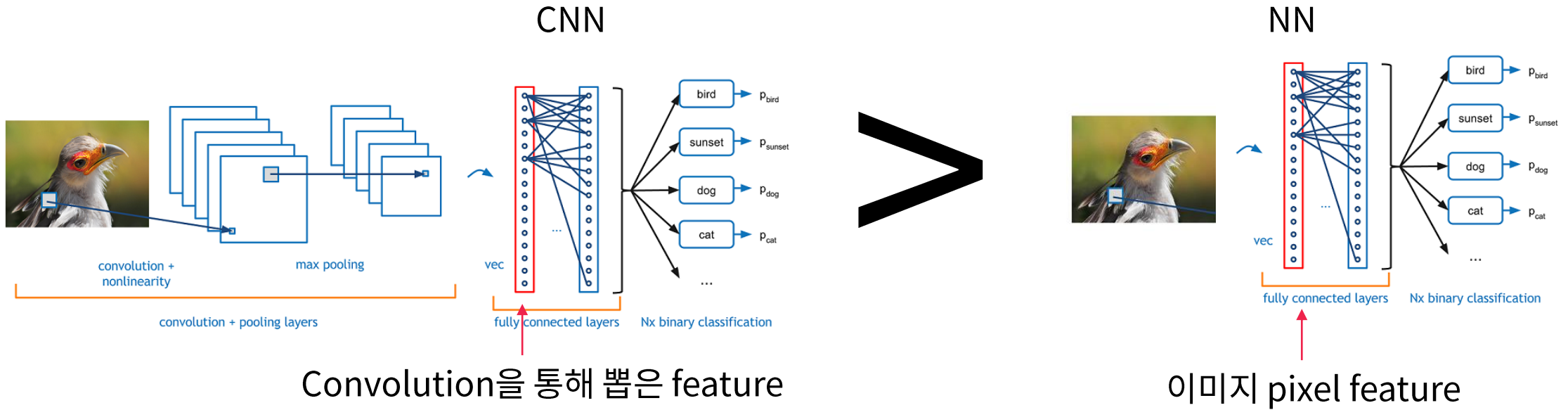
이미지 pixel feature

I DS분야에서 Tree기반 모델이 쓰이는 이유

- 딥러닝 vs 머신러닝
 - 최근 10년간은 딥러닝의 부흥기임
 - 그럼에도 불구하고 kaggle대회를 비롯한 각종 머신러닝 대회에서는 boosting계열 알고리즘이 우승함
 - 그 이유는?

IDS분야에서 Tree기반 모델이 쓰이는 이유

- 딥러닝 vs 머신러닝
 - 이미지의 지역별 feature를 뽑아서 neural network학습
 - CNN의 해주는 역할은 이미지의 feature를 잘 뽑기 위한 것 뿐



- 딥러닝은 Graphical representation을 잘 하기 위해 발전되어 옴

I DS분야에서 Tree기반 모델이 쓰이는 이유

■ 딥러닝 vs 머신러닝

- 게임유저 이탈 예측 문제 – graphical feature가 있나?

Userid	게임누적접속일	비용대비 누적획득 경험치	비용대비 평균획 득경험치	비용대비 누적던 전 클리어비율	...	이탈여부
1	4	46	11.5	0.185		잔존
2	2	180	90	0.466		이탈
			...			이탈

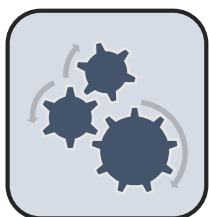
- Graphical representation learning??

IDS분야에서 Tree기반 모델이 쓰이는 이유

■ 딥러닝 vs 머신러닝

- 게임유저 이탈 예측 문제 – 학습하는 원리를 생각해보자

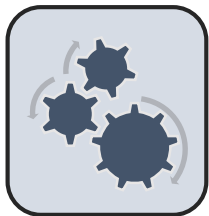
Userid	게임누적접속일	비용대비 누적획득 경험치	비용대비 평균획 득경험치	비용대비 누적던 전 클리어비율	...	이탈여부
1	4	46	11.5	0.185		잔존
2	2	180	90	0.466		이탈
			...			이탈



Tree

- 비용대비 누적 획득 경험치가 낮으면 이탈
- 비용대비 누적 던전 클리어 비율이 낮으면 이탈

...



NN

- 수많은 직선의 결합으로 비선형적인 관계를 모델링하여 잔존과 이탈을 분리

I DS분야에서 Tree기반 모델이 쓰이는 이유

- 딥러닝 vs 머신러닝

- 기존 머신러닝 모델 : 기본적으로 독립변수들은 서로 독립이라고 가정
- NeuralNetwork : Input feature들의 연속성을 가정함 (Image와 같이)

Part.05

Clustering

| Clustering이란

FASTCAMPUS
ONLINE

머신러닝과 데이터분석 A-Z

강사. 이경택