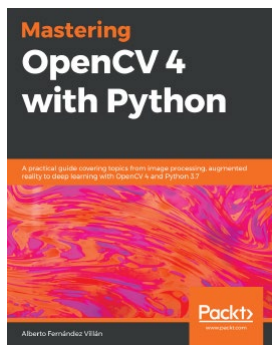




디지털영상처리/컴퓨터비전

Machine Learning

10장 내용 개념 중심으로..



Mastering OpenCV 4
with Python, Alberto,
Packt, Pub. 2019

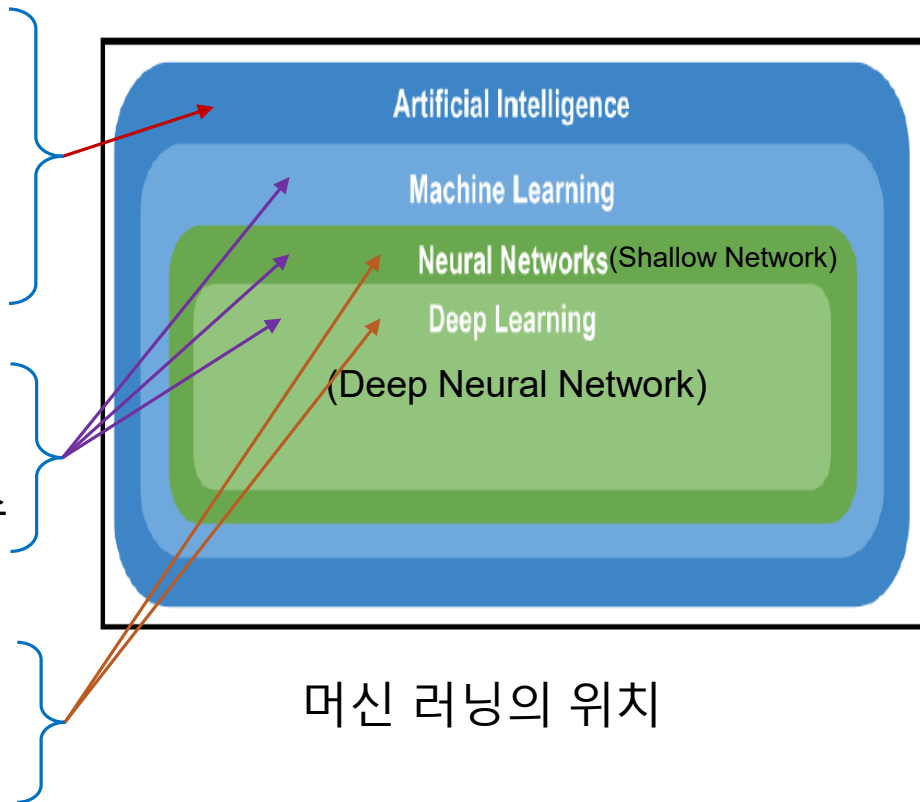
2021년 2학기

서경대학교 김진헌

머신 러닝의 개념

1

- 머신러닝의 특징: 데이터로부터 학습하는 기능이 있다. 모델로 만든다.
 - 수학적인(statistics, stochastic process, matrix theory) 모델링을 주로 사용
 - K-nearest neighbor, K-means clustering, SVM, HMM, PCA, ICA
- rule based approach
 - 규칙을 수립하고 이를 기반으로 새로운 규칙을 만들어 내는 추론 엔진(Knowledge based inference engine)
 - 적용 사례: 전문가 시스템(Expert System)
 - Language: LISP, PROLOG
- data driven approach
 - 규칙을 찾아주어 공급하지 않는다.
 - 학습 과정이 존재: 취득한 데이터로부터 스스로 모델링한다
- Artificial(초창기) Neural Network
 - 신경세포(neuron)의 기능과 그것들의 연결 구조를 수학적으로 모델링 ⇒ 뇌와 유사한 동작 구현



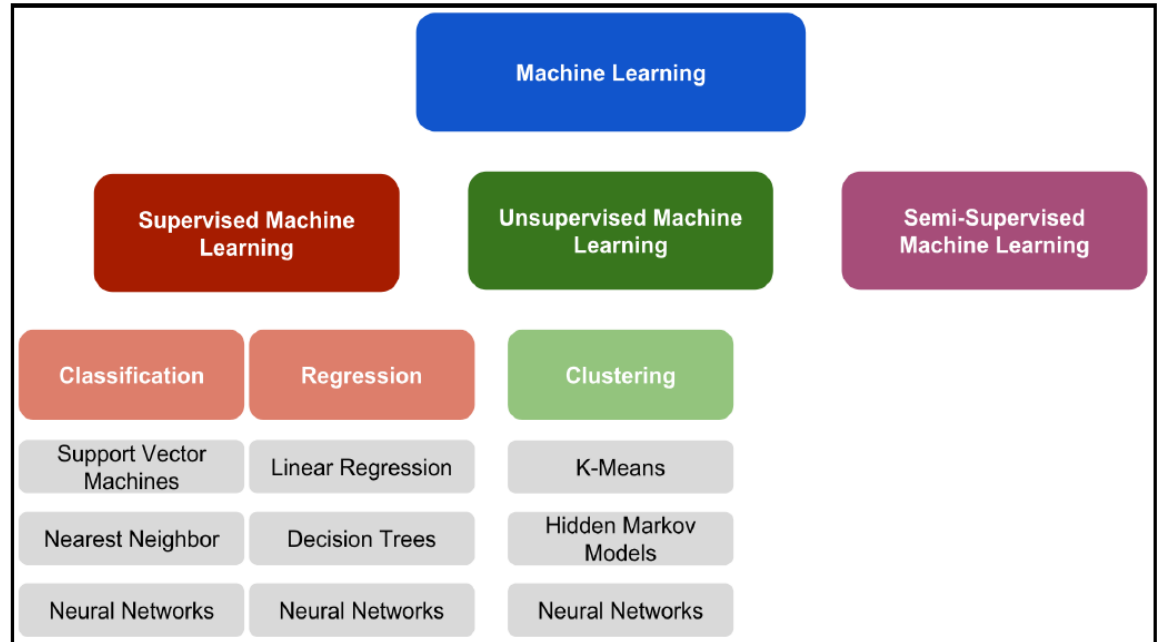
머신러닝 분야의 기술

2

Classification: 수 많은 입력이 N개 중의 어느 하나로 맵핑하도록 모델링해 놓고, 미지의 입력이 N개 중의 어디에 속하는가?

Regression: 수 많은 입력에 대해 적정한 값(아날로그)을 출력하도록 학습된 모델에 대해 새 입력은 얼마의 값으로 근사화할 수 있겠는가?

Clustering: 분류에 활용할 모델 수립의 문제를 다룬다.



Supervised Learning: 모델링하는데 이런 입력이면 이런 출력이 나와야 한다는 원칙으로 모델을 학습시킴

Unsupervised Learning: 어떤 입력에 대해 바람직한 출력이 어떤 것인지 지정하지 않고 스스로 그런 결정이 이루어지도록 모델링되는 학습 기법

Machine learning models cheat sheet

Supervised learning	Unsupervised learning	Semi-supervised learning	Reinforcement learning
<p>Data scientists provide input, output and feedback to build model (as the definition)</p> <p>EXAMPLE ALGORITHMS:</p> <p>Linear regressions</p> <ul style="list-style-type: none"> ■ sales forecasting ■ risk assessment <p>Support vector machines</p> <ul style="list-style-type: none"> ■ image classification ■ financial performance comparison <p>Decision tree</p> <ul style="list-style-type: none"> ■ predictive analytics ■ pricing 	<p>Use deep learning to arrive at conclusions and patterns through unlabeled training data.</p> <p>EXAMPLE ALGORITHMS:</p> <p>Apriori</p> <ul style="list-style-type: none"> ■ sales functions ■ word associations ■ searcher <p>K-means clustering</p> <ul style="list-style-type: none"> ■ performance monitoring ■ searcher intent 	<p>Builds a model through a mix of labeled and unlabeled data, a set of categories, suggestions and exemplar labels.</p> <p>EXAMPLE ALGORITHMS:</p> <p>Generative adversarial networks</p> <ul style="list-style-type: none"> ■ audio and video manipulation ■ data creation <p>Self-trained Naïve Bayes classifier</p> <ul style="list-style-type: none"> ■ natural language processing 	<p>Self-interpreting but based on a system of rewards and punishments learned through trial and error, seeking maximum reward.</p> <p>EXAMPLE ALGORITHMS:</p> <p>Q-learning</p> <ul style="list-style-type: none"> ■ policy creation ■ consumption reduction <p>Model-based value estimation</p> <ul style="list-style-type: none"> ■ linear tasks ■ estimating parameters

Supervised Learning

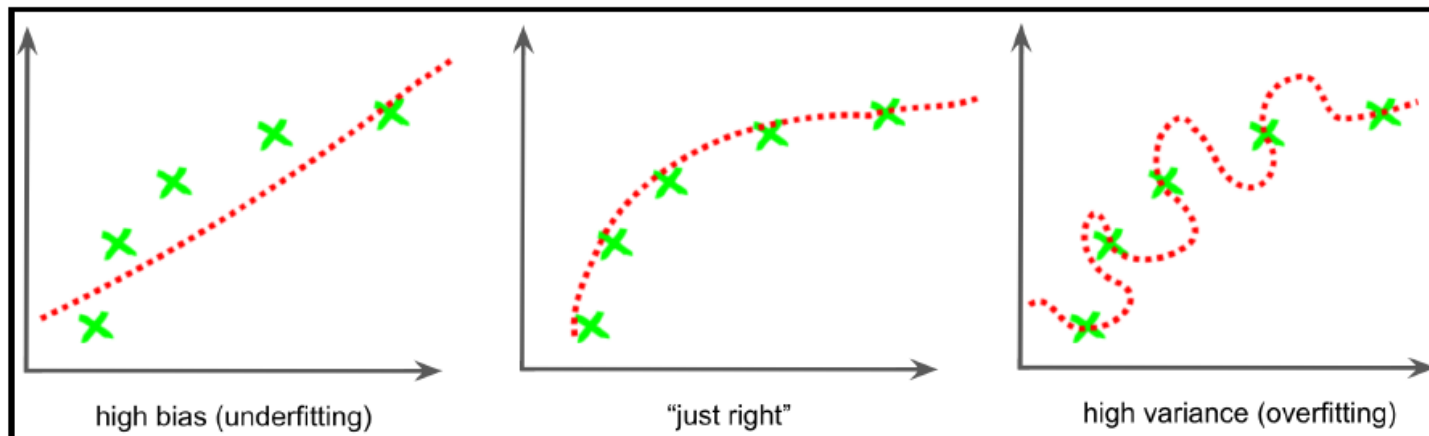
4

- 입력 데이터에 대해 어떤 결과를 출력해야하는지 알려주면서 학습 과정을 진행
- 학습과정에서 오류를 최소화하도록 학습 모델 내부의 파라미터를 조정하는 작업을 반복
- 목표가 2가지로 그룹핑된다.
 - ▣ Classification
 - 입력 데이터를 카테고리화하여 레이블링한다. 출력이 레이블의 번호이다.
 - 예: 문자의 이진 영상을 학습시켜 문자 A, 문자 B.. 영상 데이터를 학습시킨다.
 - 임의의 문자 이진 영상에 대해 문자 A(100), 문자 B(010), 문자 C(001) 레이블을 출력한다.
 - ▣ Regression
 - 출력이 연속적인 값이다.
 - 예: 타원의 원주에 따라 분포한 점들을 모델링한다.
 - 점들의 분포를 입력으로 주면 반지름은 5.2이고 장축은 10, 단축은 2.5인 것을 산출한다.

Issues of Supervised Learning

5

- Training vs Testing= 학습 vs 테스트
 - ▣ Training: 학습데이터를 사용하여 모델을 만족할 수준까지 내부의 파라미터를 안정화하는 과정
 - ▣ Testing: 학습에 사용하지 않는 데이터(test data)를 사용해서 학습의 완성도를 점진하거나 실전에 배치하는 작업이다.
- Overfitting vs Underfitting
 - ▣ Overfitting: Generalization 실패. 학습하지 않은 데이터에 대해 모두 거절하는 현상. 모델에 큰 Variance가 존재. High Variance.
 - 예: 학습시키지 않는 유사한 형태 혹은 잡음이 일부 들어간 문자 A를 A라고 인정하지 않는다.
 - ▣ Underfitting: 학습 데이터조차도 학습이 이루어지지 않는 현상. 모델에 bias가 존재. High Bias



녹색 점은 학습데이터, 붉은 선은 모델링된 결과이다. 녹색 점들 사이의 테스트 데이터가 존재한다고 가정하면 근사화되지 못한 모습을 상상할 수 있다.

Issues of Supervised Learning

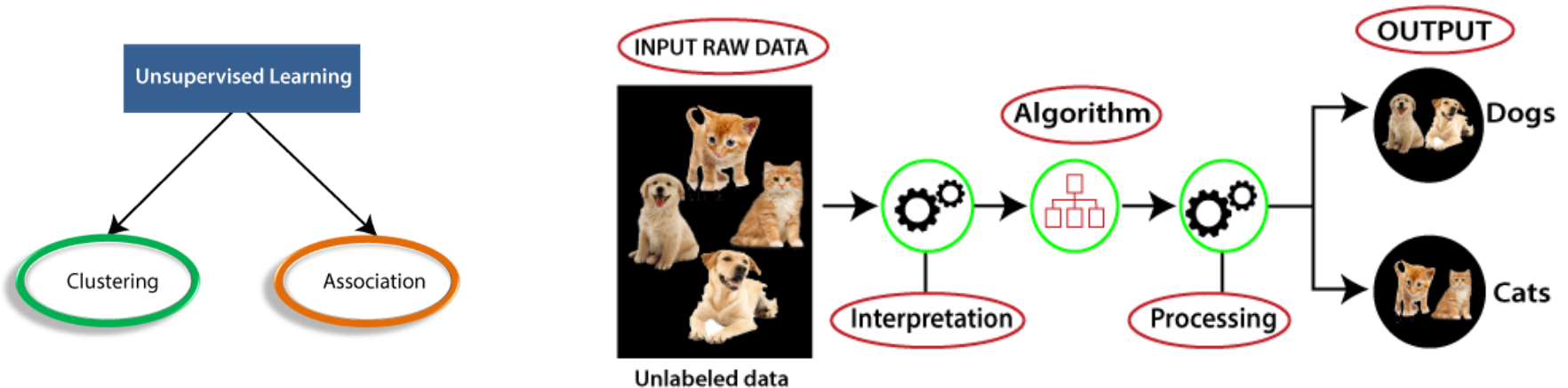
6

- Function complexity and the amount of training data
 - ▣ 트레이닝 데이터의 복잡도가 높으면 모델이 커야 한다. 데이터량이 적으면 모델이 적은 것이 바람직하다.
 - ▣ A high-complexity model will overfit the small training set.
 - ▣ 걱정 지점을 찾는 것이 어려울 때가 있다. \Rightarrow 100% 만족을 달성하기 어렵다.
- Dimensionality of the input space
 - ▣ 입력의 차원이 높으면 모델이 high variance를 갖게되어 fitting이 잘 안된다. \Rightarrow low variance를 갖는 알고리즘의 개발이 필요
 - ▣ 과거에는 입력 신호의 차원을 줄이기 위해 feature engineering을 시행
 - 모멘트 계산, 히스토그램 분석, 푸리에 변환...
 - ▣ DNN에 들어서는 대용량 데이터와 deep network가 문제를 해결하거나 or CNN 처럼 특징을 추출하는 네트워크를 앞단에 삽입하여 해결하는 상황
- Noise in the output values
 - ▣ 학습전에 잡음 학습 영상을 검출하고 제거
 - ▣ 오버 피팅을 방지하기 전에 미리 중지(early stopping)

Unsupervised Learning

7

- 출력값이 레이블이 아니라 그 자체이다. → 활용할 때는 레이블링할 수도 있다.
- 학습 데이터에 대한 레이블이 없다.
- 비지도학습의 목표는 데이터의 분포와 내재된 구조를 모델링하고 추론하는 것이다.
- 비지도학습은 올바른 응답을 찾지 않는다. → 뭐가 올바른지 정의할 수 없는 것.

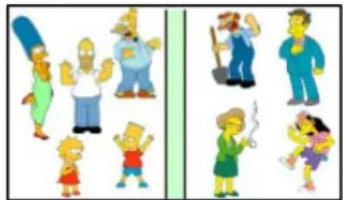


이런 문제는 비지도학습에 적당치 않은 사례이지만, 이해를 돕기 위해 이 문제를 비지도 학습으로 푼다면 이런 경로를 취할 것이다하는 정도로 이해해 주기 바람.

clustering

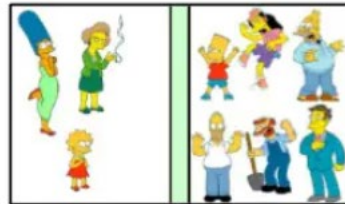
8

What is natural grouping among these objects?



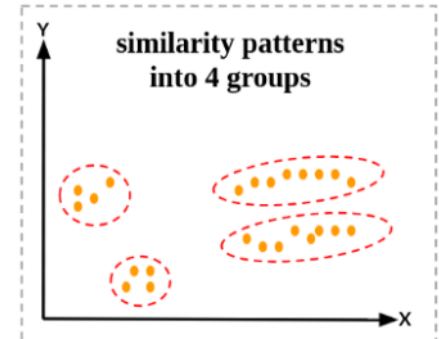
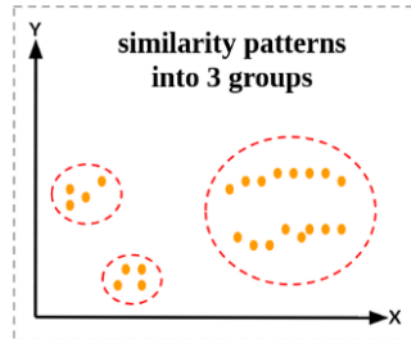
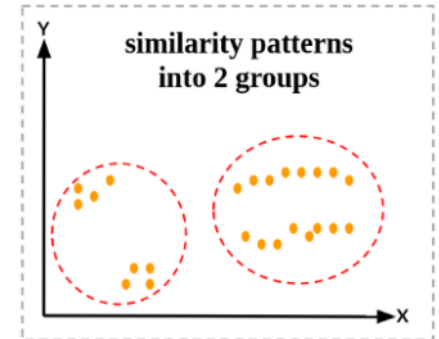
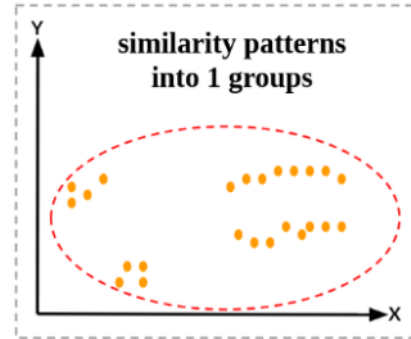
Simpson's Family

None Simpson's Family
(School Employees)



Females

males



참고링크:OpenCV가 다루고 있는 머신 러닝 자료 소개(1)

- in python [opencv machine learning tutorials](#)

9

- 위치: <http://docs.opencv.org> ⇒ main page ⇒ opencv-python tutorials ⇒ machine learning ⇒ [opencv machine learning tutorials](#)



• K-Nearest Neighbour

- Learn to use kNN for classification Plus learn about handwritten digit recognition using kNN

• Support Vector Machines (SVM)

- Understand concepts of SVM



• K-Means Clustering

- Learn to use K-Means Clustering to group data to a number of clusters. Plus learn to do color quantization using K-Means Clustering

참고 링크: OpenCV가 다루고 있는 머신 러닝 자료 소개(2)

- in C++, Java, python ML module tutorial

10

- 위치: <http://docs.opencv.org> ⇒ main page ⇒ opencv python tutorials ⇒ machine learning(ml module)

Use the powerful machine learning classes for statistical classification, regression and clustering of data.

Learn what a Support Vector Machine is.

- Introduction to Support Vector Machines

Here you will learn how to define the optimization problem for SVMs when it is not possible to separate linearly the training data.

- Support Vector Machines for Non-Linearly Separable Data

Learn what a Principal Component Analysis (PCA) is.

- Introduction to Principal Component Analysis (PCA)

참고: Classes of OpenCV ml module(1)

11

class	cv::ml::ANN_MLP
	Artificial Neural Networks - Multi-Layer Perceptrons. More...
class	<u>cv::ml::Boost</u>
	Boosted tree classifier derived from DTrees . More...
class	cv::ml::DTrees
	The class represents a single decision tree or a collection of decision trees. More...
class	cv::ml::EM
	The class implements the Expectation Maximization algorithm. More...
class	cv::ml::KNearest
	The class implements K-Nearest Neighbors model. More...
class	cv::ml::LogisticRegression
	Implements Logistic Regression classifier. More...
class	cv::ml::NormalBayesClassifier
	Bayes classifier for normally distributed data. More...

참고: Classes of OpenCV ml module(2)

12

class **cv::ml::ParamGrid**

The structure represents the logarithmic grid range of statmodel parameters. [More...](#)

class **cv::ml::RTrees**

The class implements the random forest predictor. [More...](#)

struct **cv::ml::SimulatedAnnealingSolverSystem**

This class declares example interface for system state used in simulated annealing optimization.

class **cv::ml::StatModel**

Base class for statistical models in OpenCV ML. [More...](#)

class **cv::ml::SVM**

Support Vector Machines. [More...](#)

class **cv::ml::SVMMSGD**

Stochastic Gradient Descent **SVM** classifier. [More...](#)