

# Class Summary

---

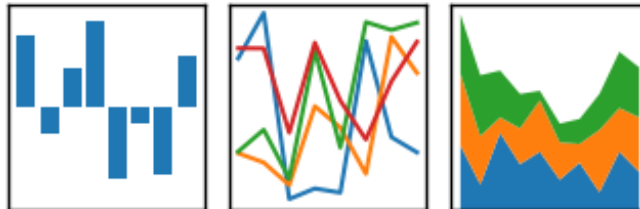
Kyungsik Han

# 학습 내용

- 머신 러닝, 데이터 사이언스에 대한 이해
- 개발 환경 구축 (Anaconda)
- 파이썬에 대한 이해
  - 데이터베이스
  - Numpy
  - Pandas (dataframe, series)

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



# 학습 내용

- 다양한 머신러닝 알고리즘에 대한 학습
  - 지도 학습
  - 비지도 학습
- 지도 학습
  - 선형 (regression) 분석
    - 회귀 분석
  - 분류 (classification) 분석
    - Logistic Regression
    - Support Vector Machine
    - Decision Tree
    - Random Forest
    - k-Nearest Neighbor

# 학습 내용

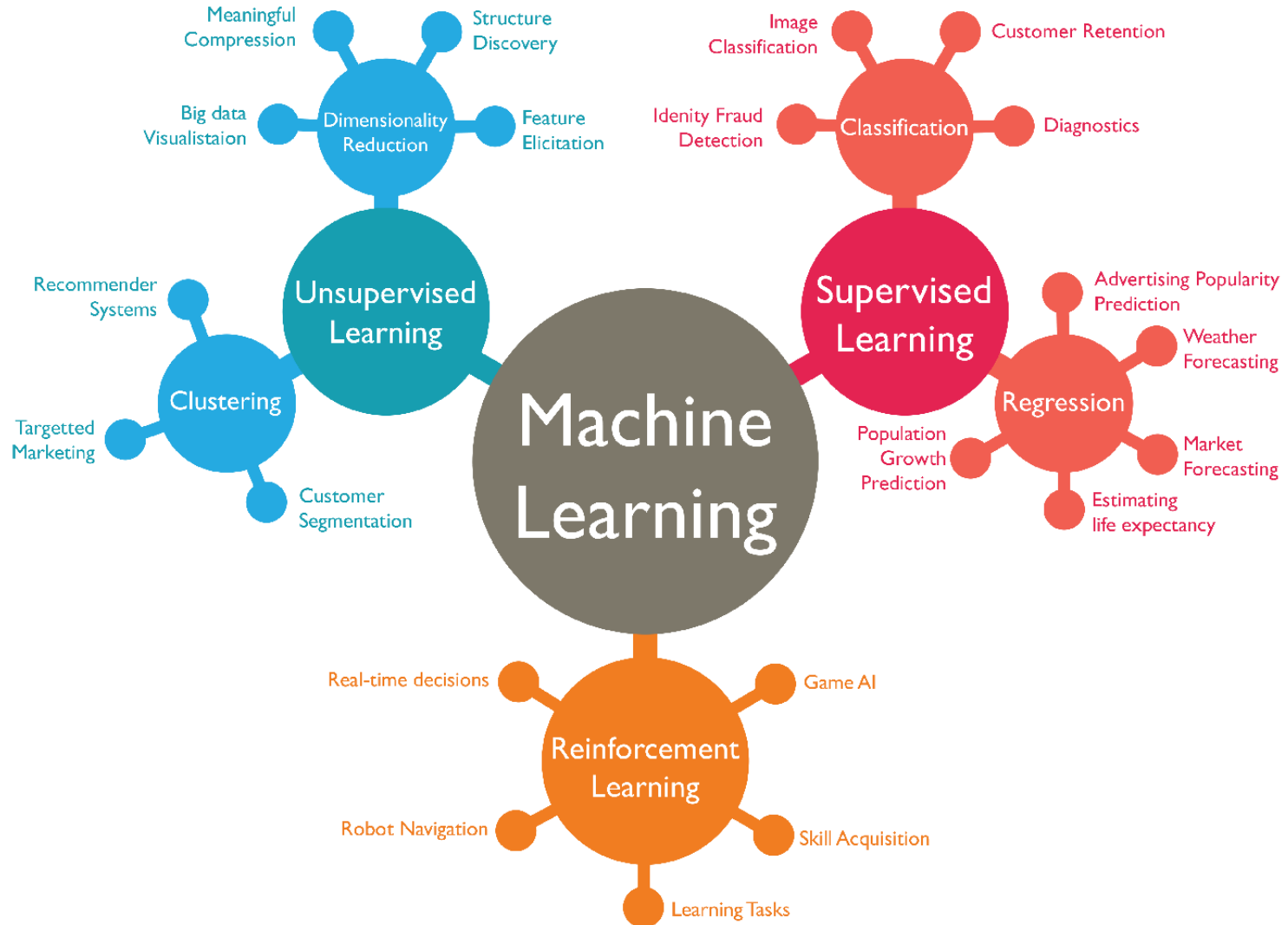
- 비지도 학습
  - k-Means Clustering
  - Hierarchical Clustering
- 데이터 전처리
  - Scaling: Standard Scaler, MinMaxScaler, ...
- 모델 평가
  - 훈련데이터 & 테스트 데이터
  - 과적합 (Overfitting)
  - 교차 검증 (Cross Validation)
  - 그리드서치 (Grid Search) ➔ 차후 학습 필요

# 학습 내용

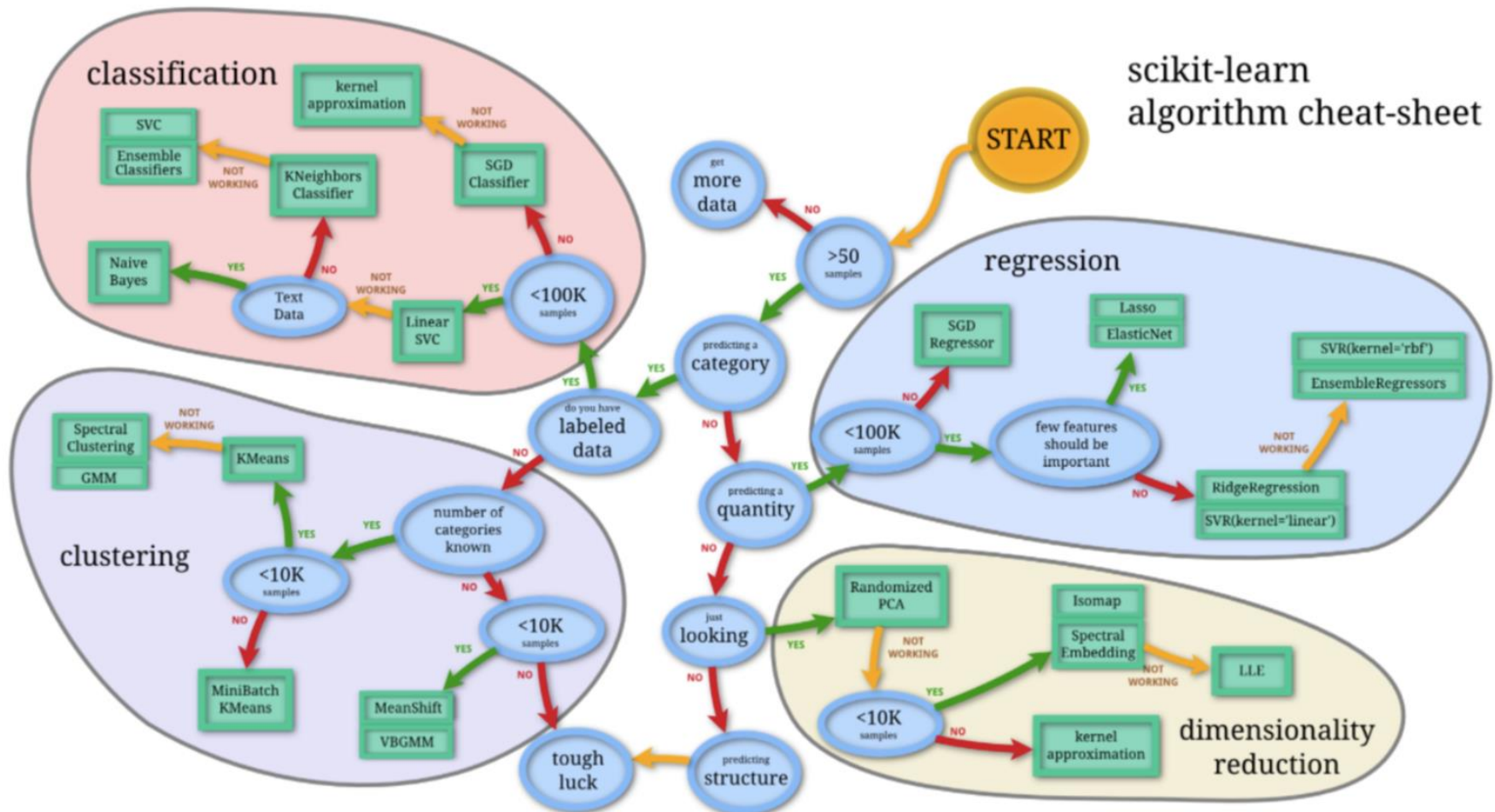
- 분류 성능 평가
  - Accuracy, Precision (정밀도), Recall (재현율), F1-score
  - ROC (Receiver Operating Characteristics), AUC
- 회귀 성능 평가:  $R^2$

- Feature Selection
  - 특성 파악을 위한 다양한 방법 학습
    - 일변량 통계 (univariate statistics)
    - 모델 기반 선택 (model-based selection)
    - 반복적 선택 (iterative selection)
    - 차원 축소 (dimensionality reduction)

# 이해가 되는지?

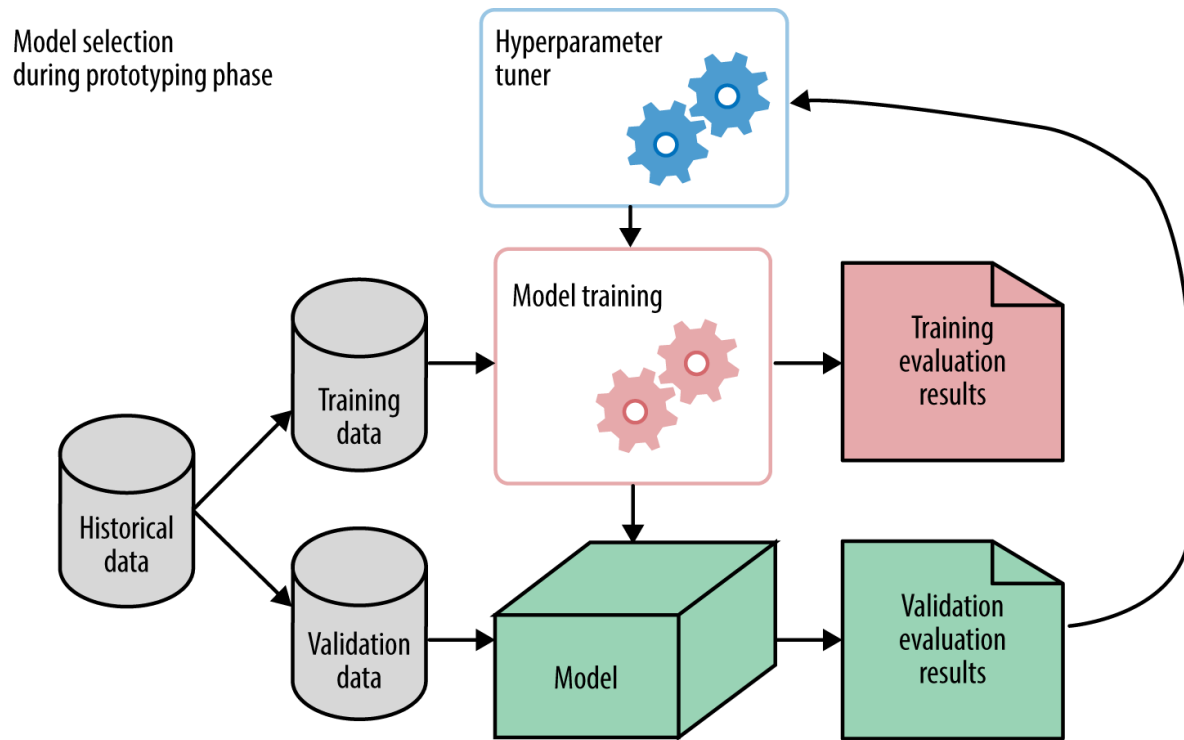
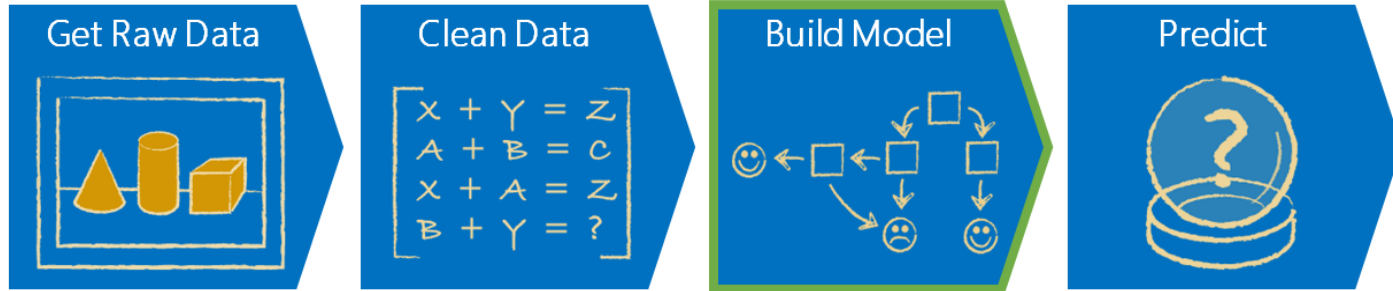


# 이해가 되는지?





# 이해가 되는지?



# 머신러닝 활용에 대해

- 효과적인 머신러닝 사용을 위해서는 넓은 시각으로 문제를 바라보아야 함
  - 목적을 먼저 파악
    - 부정 거래 탐지?
    - 영화 추천?
  - 목적의 성공을 어떻게 정의하고 측정할지?
  - 최종 솔루션이 비즈니스 or 연구 목표에 어떤 영향을 줄지 생각
    - **So What** Question

# 머신러닝 활용에 대해

- 목적: 부정 거래 탐지
  - 부정거래 예측이 실제로 작동하는지 어떻게 측정할 수 있을까?
  - 알고리즘을 평가하기에 알맞은 데이터를 가지고 있는가?
  - 데이터는 충분히 확보할 수 있는 것인가?
  - 성공적인 모델 구축 후, 이 솔루션이 비즈니스에 어떤 영향을 주는가?

# 머신러닝 활용에 대해

- 목적: 온라인 댓글 감정 분류 모델
  - 감정의 종류는 정확하게 구분할 수 있는 요소인지?
    - 기쁨, 슬픔, 분노, 혐오 등
  - 사람마다 느끼고 판단하는 댓글의 감정은 다를 수 있는데 어떤 방식으로 감정을 라벨링(labeling) 해야 될까?
    - 감정 사전 기반으로 댓글 감정 결정?
    - 사람이 직접 판단? 다수의 결정으로 감정 결정?
  - 데이터의 크기는 얼마나 되어야 될까?

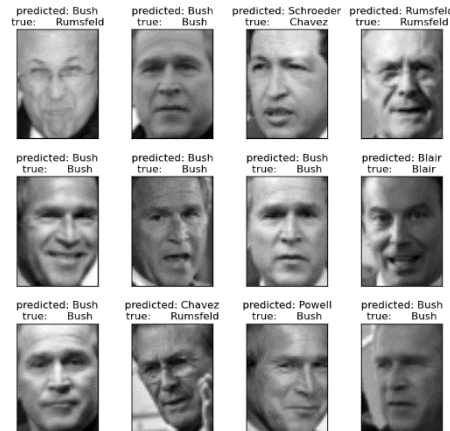
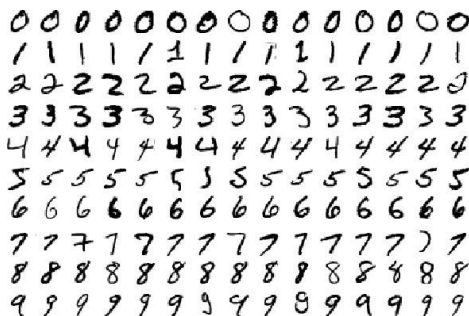
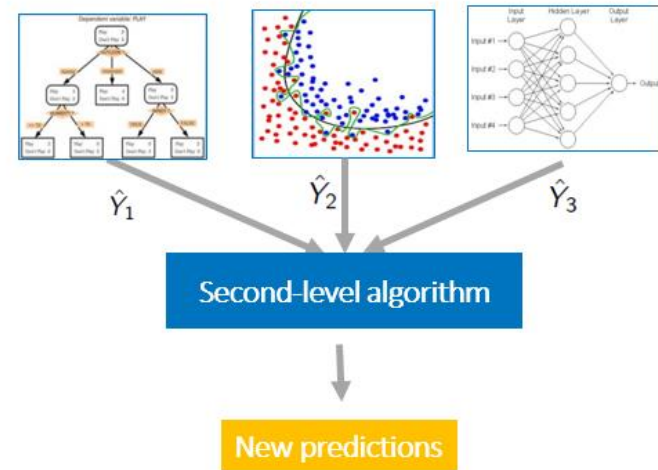
# 앞으로 해야될 것

- 이론

- 본 강의에서는 이론에 대해서 깊이 다루지 않았음
- 많은 모델이 확률, 선형 대수, 최적화 등의 이론을 사용
- 알고리즘 이면의 이론을 알아두면 실습의 과정에 대한 이해가 깊어질 수 있음
- 이론을 설명하고 있는 교재가 많으니 참고하는 것을 추천

# 앞으로 해야될 것

- 아직 학습해야 될 머신러닝 종류는 많음
  - Other models: Naïve Bayes, Many other ensemble methods
  - Some popular examples
    - MNIST
    - Face recognition



# 앞으로 해야될 것

- 활용 예제
  - 머신 러닝의 다양한 활용 사례를 참고하면서 이해도를 높임
  - 시중에 많은 교재가 있으니 참고하는 것을 추천
- 신경망
  - Deep Learning
  - Python Package (Tensorflow, Keras)가 많이 있기 때문에 앞으로 학습과 활용에 대한 고민도 필요
  - 꾸준히 책도 출판되고 있음

- Machine learning 구현에 어느정도 익숙해졌다면 deep learning 도 시작해볼 수 있다
  - Keras, Tensorflow, PyTorch
  - DNN (Deep Neural Network)
  - CNN (Convolutional Neural Network)
  - RNN (Recurrent Neural Network)
  - LSTM (Long Short-Term Memory)
  - AutoEncoder
  - Sequence-to-sequence (NLP)



수고하셨습니다