

Chapter 04 Classifaction

4 분류를 해보자!
1편

2020 Ver.

D.COM
DATA-SCIENCE

이전까지
데이터를 정제하고 시각화 하는 방법에 대해
알아보았습니다.

이번 시간부터는
데이터를 활용하여
예측(Prediction)
하는 방법에 대해 알아보니다.

예측(Prediction)

파악된 관계를 이용하여
우리가 원하는
새로운 데이터를 만들어 내는 과정

Q. 지금 받은 메일이 스팸 메일일까 아닐까?

Q. 이러한 프로필을 가진 사용자가 가장 좋아하는 노래 장르는 무엇일까?

Q. 이 사진은 고양이 사진일까 강아지 사진 일까?

Q. 강남 어떤 지역의 30평 대 아파트의 가격은 대략 얼마일까?

=> 예측 문제의 예

과거에는 어떤 데이터를 보고
예측하기 위해 규칙을
사람이 일일이 작성하였습니다.

이를 규칙 기반 학습(Rule-Based-Learning)

이라고 합니다.



하지만 다음 사진을 보고
사과임을 예측할 수 있는 모델은
어떻게 제작할 수 있을까요?

사람이 일일이 규칙을 작성하기에는
너무 복잡하고, 예외가 많습니다.

따라서 제안된 방법이 바로
기계 학습(Machine Learning)입니다.

머신러닝 VS 딥러닝 VS 인공지능

세 가지 개념을 확실히 잡아봅시다!

링크

지도 학습 VS 비지도 학습 VS 강화 학습

세 가지 개념을 확실히 잡아봅시다!

링크

이번 단원은,

지도 학습(Supervised Learning) 중,
이진 분류(Binary Classification)에 대해 배웁니다.

대표적인 이진분류 알고리즘 4가지를 배워봅시다!

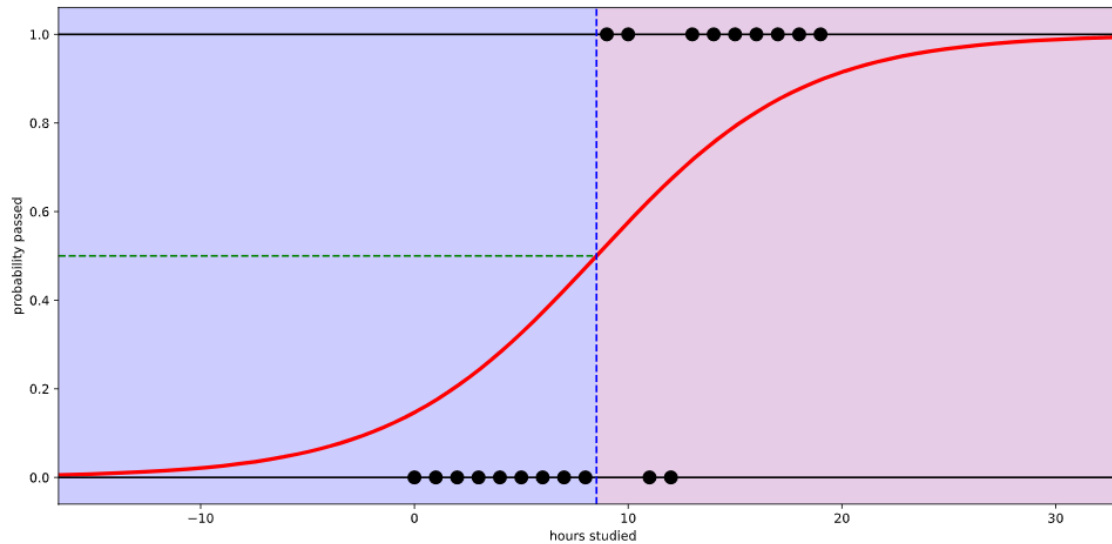
- 1. Logistic Regression**
- 2. SVM (Support Vector Machine)**
- 3. Decision Tree**
- 4. KNN (K-Nearest Neighbor)**

분류 알고리즘의 핵심은
분류하는 ‘기준’을 생성하는 것입니다.

4가지 알고리즘이 분류하는 기준을 어떻게 세우는지 알아보시다.

각 알고리즘의 구체적인 구현이나 수식은
일단은 넘어 가셔도 좋습니다!
우선은 동작 개념에 집중해주세요!

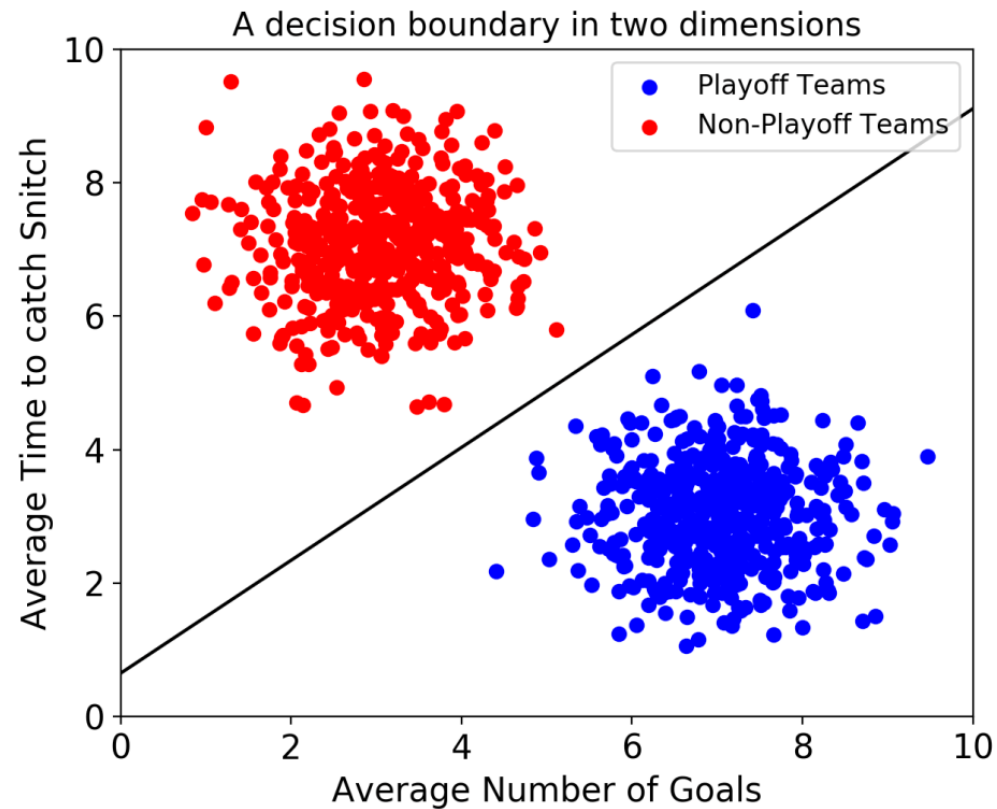
Logistic Regression



회귀(Regression)과 유사한 방법으로
분류를 위한 모델 생성

<http://hleecaster.com/ml-logistic-regression-concept/>

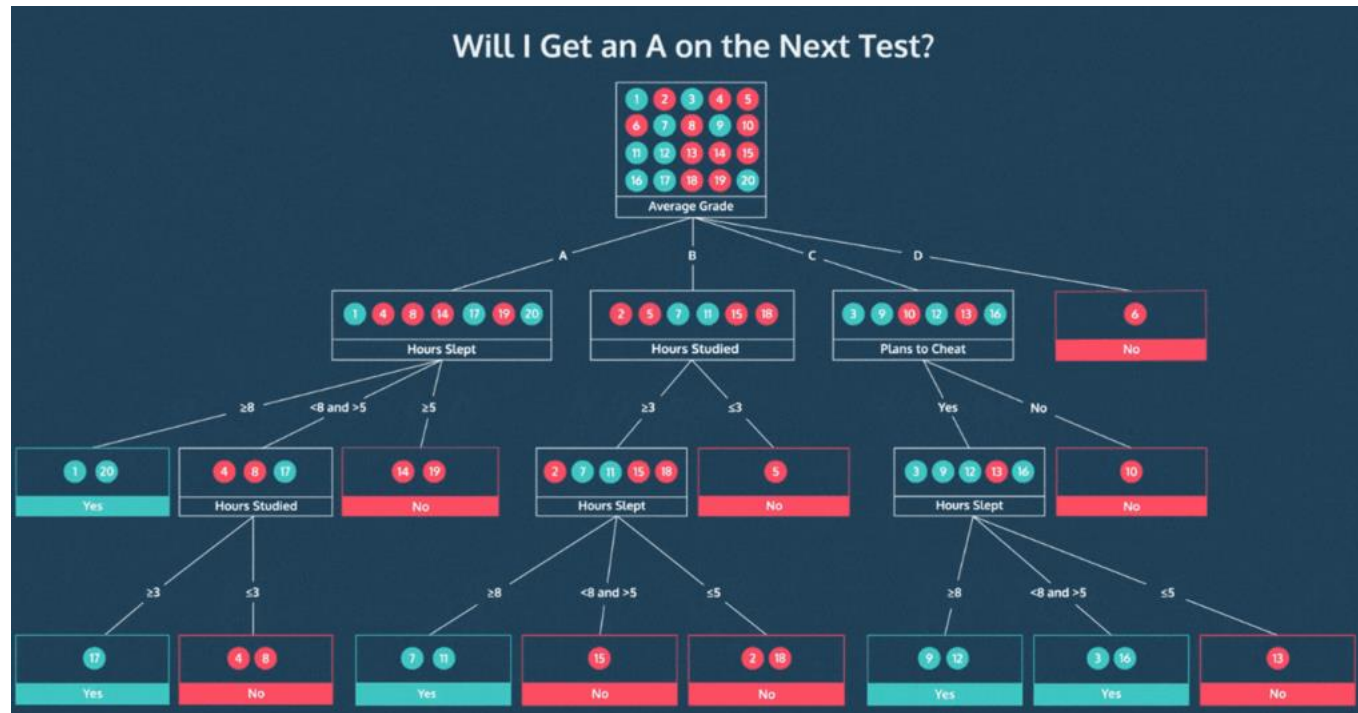
Support Vector Machine



분류를 위한
경계를 그려보자!

<http://hleecaster.com/ml-svm-concept/>

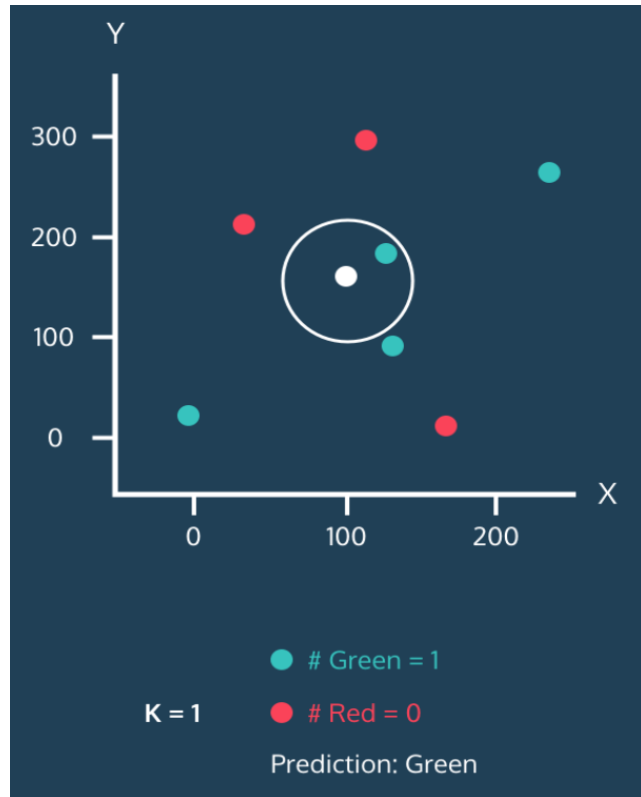
Decision Tree



분류를 위한
체크 리스트를 만들어보자!

<http://hleecaster.com/ml-decision-tree-concept/>

KNN



비슷한 데이터로
묶어서 분류해보자!

<http://hleecaster.com/ml-knn-concept/>

Scikit-learn

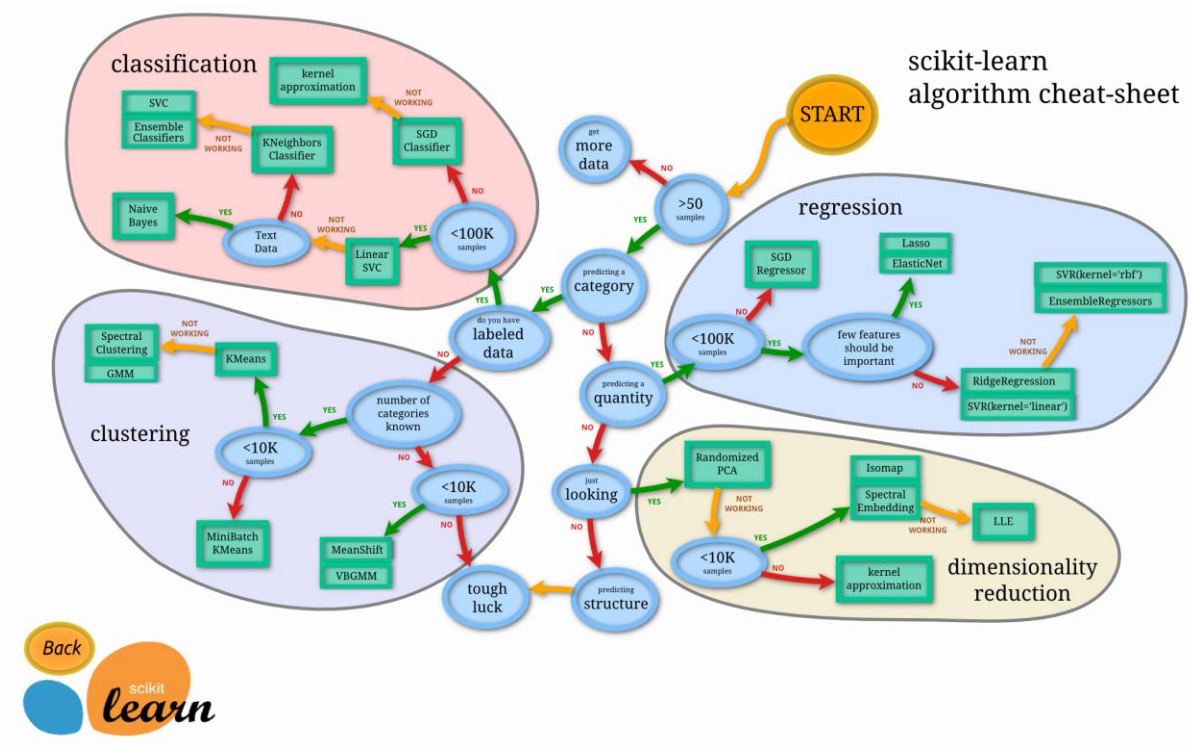
라이브러리를 활용하면
앞선 머신러닝 알고리즘들을
매우 손쉽게 구현할 수 있습니다.

파이썬 사이킷런 기초

<https://losskatsu.github.io/machine-learning/sklearn/#%EA%B0%80%EC%83%81-%EB%8D%B0%EC%9D%B4%ED%84%B0-%EB%9E%9C%EB%8D%A4%EC%9C%BC%EB%A1%9C-%EC%83%9D%EC%84%B1%ED%95%98%EA%B8%B0>

사이킷런으로 붓꽃 품종 예측하기

<https://hun931018.tistory.com/27>



어떤 머신러닝 알고리즘을
선택해야 할까요?
각 알고리즘에는 장단점이 있습니다!

https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html

이번 Practice에서는
Scikit-Learn을 이용하여
예측 모델을 제작해 봅시다.