



# Machine Learning Model

Titanic 생존자 예측



## 목 차

- ✓ Logistic Regression
- ✓ SVM(Support Vector Machines)
- ✓ K-NN(K Nearest Neighborhood)
- ✓ Gaussian Naive Bayes
- ✓ Perceptron
- ✓ Linear SVC
- ✓ Stochastic Gradient Descent(SGD)
- ✓ Decision Tree
- ✓ Random Forest



## Logistic Regression(로지스틱 회귀)

### ✓ Logistic Regression

- 독립 변수의 선형 결합을 이용하여 사건의 발생 가능성을 예측하는데 사용되는 통계 기법
- Logistic Regression의 목적은 일반 회귀 분석의 목표와 동일하게 종속 변수와 독립 변수 간의 관계를 구체적인 함수로 나타내 향후 예측 모델에 사용하는 것
- Linear Regression과 달리 종속 변수가 범주형 데이터를 대상으로 하며 입력 데이터가 주어졌을 때 해당 데이터의 결과가 특정 분류로 나뉨



## SVM(Support Vector Machines)

- ✓ SVM(Support Vector Machines)
  - SVM은 **결정 경계(분류를 위한 기준선)를 정의하는** 모델
  - 분류되지 않은 새로운 점이 나타나면 경계의 어느 쪽에 속하는지 확인해서 분류 과제를 수행할 수 있게 함



## K-NN(K Nearest Neighborhood)

- ✓ K-NN(K Nearest Neighborhood)
  - K-NN 알고리즘은 데이터로부터 거리가 가까운 'k'개의 다른 데이터의 레이블을 참조하여 분류하는 알고리즘으로 거리를 측정할 때 유클리디안 거리 계산법을 사용함
  - 이미지 처리, 영상에서 글자 인식과 얼굴 인식, 영화나 음악, 상품 추천에 대한 개인별 선호 예측, 의료, 유전자 데이터의 패턴 인식 등 많은 분야에서 응용되어 사용됨

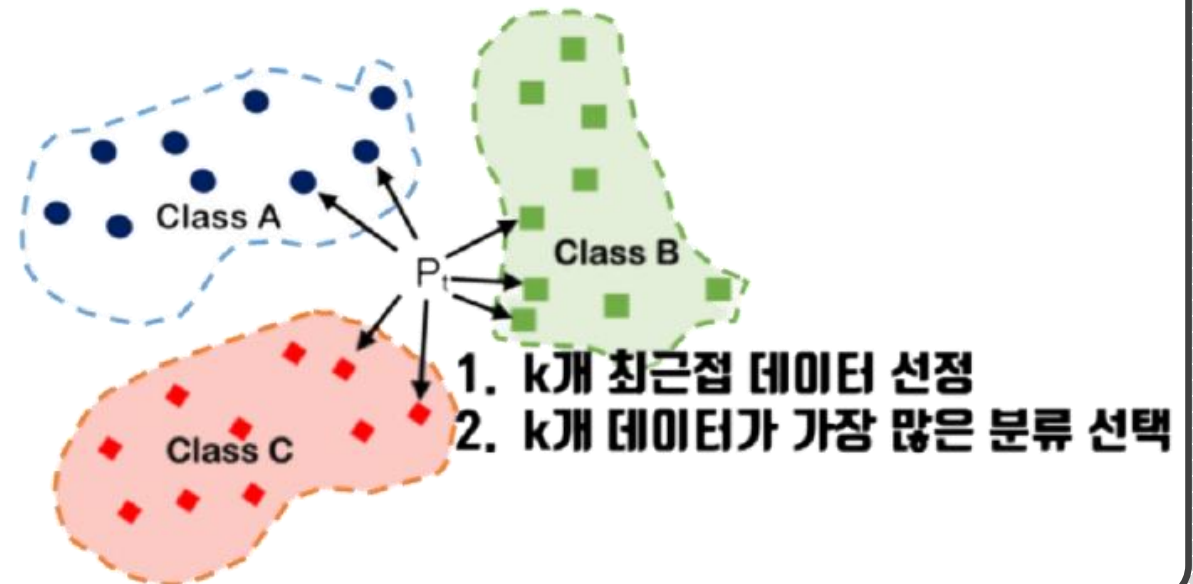


# K-NN(K Nearest Neighborhood)

## ✓ K-NN(K Nearest Neighborhood)

1. 새로운 데이터가 입력되었을 때, 기존의 데이터와 새로운 데이터를 비교함
2. 새로운 데이터와 가장 인접한 데이터 k개를 선정
3. k값에 의해 결정된 분류를 입력된 데이터의 분류로 확정

새로 입력된 데이터와 기존 데이터를  
비교함으로써 새로운 데이터를  
유사하게 판단된 기존 데이터로 분류





# Gaussian Naive Bayes

## ✓ Naive Bayesian Classification

- **각 특성을 개별(독립)로 취급해 파라미터를 학습하고 각 특성에서 클래스별 통계를 단순히 취합**
- 스팸 메일 필터, 텍스트 분류, 감정 분석, 추천 시스템 등에 광범위하게 활용되는 분류 기법
- Logistic Regression이나 Linear SVC 같은 선형 분류기보다 훈련 속도는 빠르지만, 일반화 성능이 조금 떨어짐



# Gaussian Naive Bayes

## ✓ Naive Bayesian Classification의 종류

- GaussianNB
  - 연속적인 데이터에 적용
  - 클래스별로 각 특성의 표준편차와 평균을 저장
- BernoulliNB
  - 이진 데이터에 적용
  - 복잡도를 조절하는 매개변수 하나를 가지고 있음
- MultinomialNB
  - 카운트 데이터에 적용
  - 클래스별로 특성의 평균을 계산
  - 복잡도를 조절하는 매개변수 하나를 가지고 있음





# Perceptron

## ✓ Perceptron

- Perceptron = Perception + Neuron
  - Perception : 무언가를 인지하는 능력
  - Neuron : 감각 입력 정보를 의미있는 정보로 바꿔주는 뇌에 있는 신경세포
- 인공신경망의 한 종류
- 동작방식
  - 각 노드의 '가중치\*입력치'를 모두 합한 값이 활성화함수에 의해 판단
  - 그 값이 임계치(보통 0)보다 크면 뉴런이 활성화되고 결과값으로 1을 출력
  - 뉴런이 활성화되지 않으면 결과값으로 -1을 출력



## Linear SVC

### ✓ Linear SVC

- 선형 분류에 특화된 SVM
- 선형 분류가 가능할 경우 계산이 빠름



## Stochastic Gradient Descent(SGD)

- ✓ Stochastic Gradient Descent(확률적 경사하강법)
  - 데이터 세트에서 **무작위로 균일하게 선택한 하나의 예를 의존하여 각 단계의 예측 경사를 계산**
  - 입력 데이터 한 개만을 사용하기 때문에 한 개의 데이터를 '벡터'로 표현하여 특정 레이어 층에 입력하는 것으로 이해할 수 있고 이는 벡터 x 행렬 연산이 됨



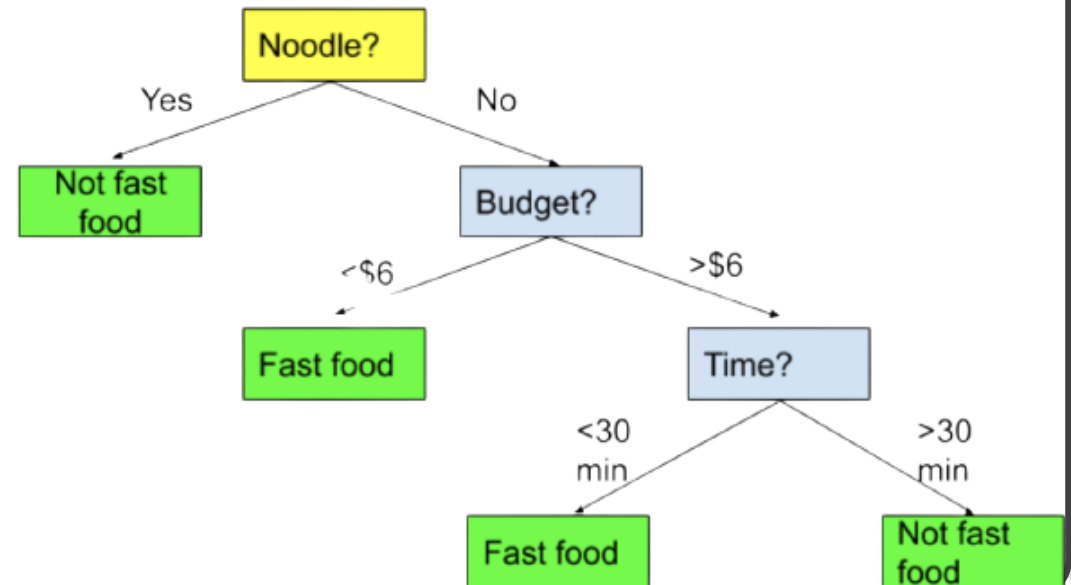
# Decision Tree(의사결정트리)

## ✓ Decision Tree

- Decision Tree는 분류와 같은 의사결정을 수행할 때, 나무와 같이 **가지치기를 하면서 분류**하는 방법
- 과거에 수집된 데이터들을 분석하여 이들 사이에 존재하는 패턴을 속성의 조합으로 나타내는 분류 모형

### • 분류절차

1. 새로운 데이터 분류
2. 해당 범주 값 예측
3. 트리 구조의 일반화된 지식 추출





## Random Forest

### ✓ Random Forest

- 훈련을 통해 구성해 놓은 **Decision Tree**들로부터 **분류 결과**를 **취합해서 예측**하는 앙상블 모델
- 여러 개의 Decision Tree를 통해 Random Forest를 만들면 오버피팅되는 단점을 해결 가능



## 참고사이트

- ✓ 템플릿: <http://pptbizcam.co.kr/?p=7576>
- ✓ <https://blog.naver.com/so15284/222127763697>
- ✓ <https://m.blog.naver.com/y4769/221851780608>
- ✓ <https://nyanye.com/til/2017/01/17/Logistic-Regression/>
- ✓ <https://sanghyu.tistory.com/7>
- ✓ <https://sacko.tistory.com/10>
- ✓ <http://hleecaster.com/ml-perceptron-concept/>
- ✓ <https://ko.wikipedia.org/wiki/%ED%8D%BC%EC%85%89%ED%8A%B8%EB%A1%A0>
- ✓ <https://bkshin.tistory.com/entry/%EB%A8%B8%EC%8B%A0%EB%9F%AC%EB%8B%9D-5-%EB%9E%9C%EB%8D%A4-%ED%8F%AC%EB%A0%88%EC%8A%A4%ED%8A%B8Random-Forest%EC%99%80-%EC%95%99%EC%83%81%EB%B8%94Ensemble>
- ✓ <https://thisisprogrammingworld.tistory.com/63>
- ✓ <https://yeo0.github.io/data/2018/10/30/5.-svm/>
- ✓ <https://everyday-deeplearning.tistory.com/entry/SGD-Stochastic-Gradient-Descent-%ED%99%95%EB%A5%A0%EC%A0%81-%EA%B2%BD%EC%82%AC%ED%95%98%EA%B0%95%EB%B2%95>