기초 이론부터 실무 실습까지

머신 러닝 익히기

programmers

Part 11. 종강

정 정 민

programmers

Chapter 25. 강의 복습 및 마무리

1. 수업에서 다룬 내용
2. RECAP

수업에서 다룬 내용

우리는 이런 영역을 다뤘습니다

지도 학습

- **분류**
  - 선형 분류
  - SVC
  - Decision Tree Regression
  - CNN
- **회귀**
  - 선형 회귀
  - SVM
  - Decision Tree

비지도 학습

- **군집화**
  - K-means
- **이상치 탐지**
  - Isolation Forest

RECAP

선형 모델

- 파라미터들이 선형 결합을 이루고,
  이것으로 종속 변수의 값을 표현할 수 있을 때
  이것을 선형 모델이라고 함

$$y = w_1 x_1 + w_2 x_2 + \cdots + w_n x_n$$

- $x_1 \ldots x_n$: 독립 변수 혹은 특징(feature), 보통 입력하는 데이터를 의미
- $w_1 \ldots w_n$: 파라미터, 찾아내야 하는 값
- y : 종속 변수

- 독립 변수가 “독립” 변수라는 이름을 갖듯 서로 다른 독립 변수는 서로 상관성이 없어야 함

선형 회귀 (Linear Regression)

- 입력 데이터 특징 사이의 독립성을 가정하고
- 데이터 특징에 대한 선형 결합으로 회귀 문제를 풀겠다는 의미

$$\hat{y} = w_0 + w_1 x_1 + w_2 x_2 + \cdots + w_p x_p = Xw$$

- 비용 함수는 MSE를 사용

$$MSE = J(w) = \|y - \hat{y}\|_2^2$$

- 최적화 방법론
  - 정규 방정식 : $w = (X^T X)^{-1} X^T y$ 혹은 $V\Sigma^{-1}U^T y$
  - 경사 하강법 : $w^{new} = w - lr \cdot \frac{\partial}{\partial w} J(w)$

선형 분류 (Logistic Regression)

- 입력 데이터 특징 사이의 독립성을 가정하고
- 선형 모델로 Logit을 예측해 확률 추정으로 각 클래스일 확률을 예측함

$$\widehat{logit} = w_0 + w_1 x_1 + w_2 x_2 + \cdots + w_n x_n = Xw$$

$$\hat{p} = \sigma\left(\widehat{logit}\right)$$

$$\hat{y} = \begin{cases} 0, & \hat{p} < 0.5 \\ 1, & \hat{p} \geq 0.5 \end{cases}$$

- 비용 함수는 로그 손실(Log Loss)을 사용

$$J(w) = -\frac{1}{N}\sum(y \cdot \log(\hat{p}) + (1-y) \cdot \log(1-\hat{p}))$$

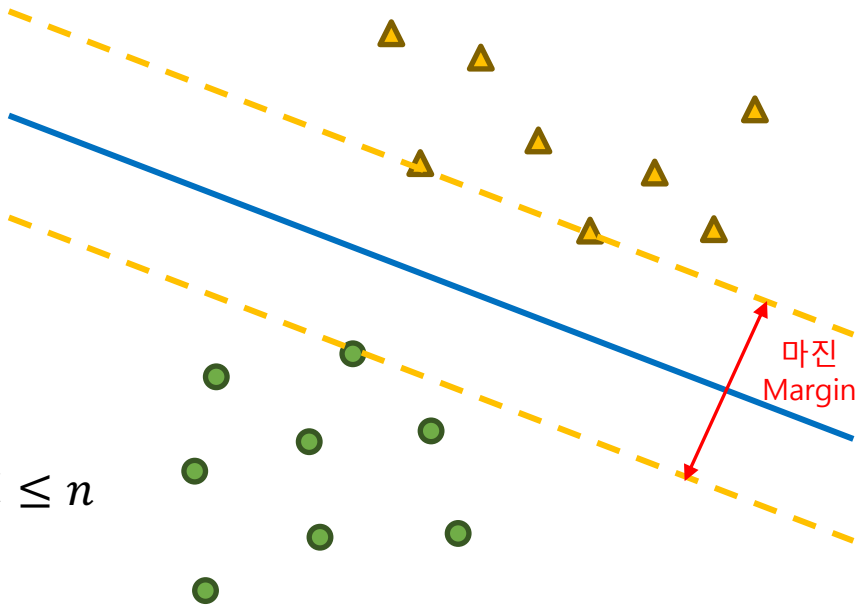
- 경사 하강법을 활용해 최적화 진행

SVM의 진행 과정과 최적화 문제

- SVM은 두 데이터 분류하는 최적의 직선(고차원의 경우 초평면) 을 찾고자 함
- 여기서 최적 직선은 마진을 최대화 하는 직선!
  - 이때의 직선을 '최대 마진 초평면'이라고 함
- 하드 마진 SVM → 소프트 마진 SVM
  - 슬랙 변수 도입
- 선형 SVM → 비선형 SVM
  - 커널 트릭 활용
- 최종 최적화 식은

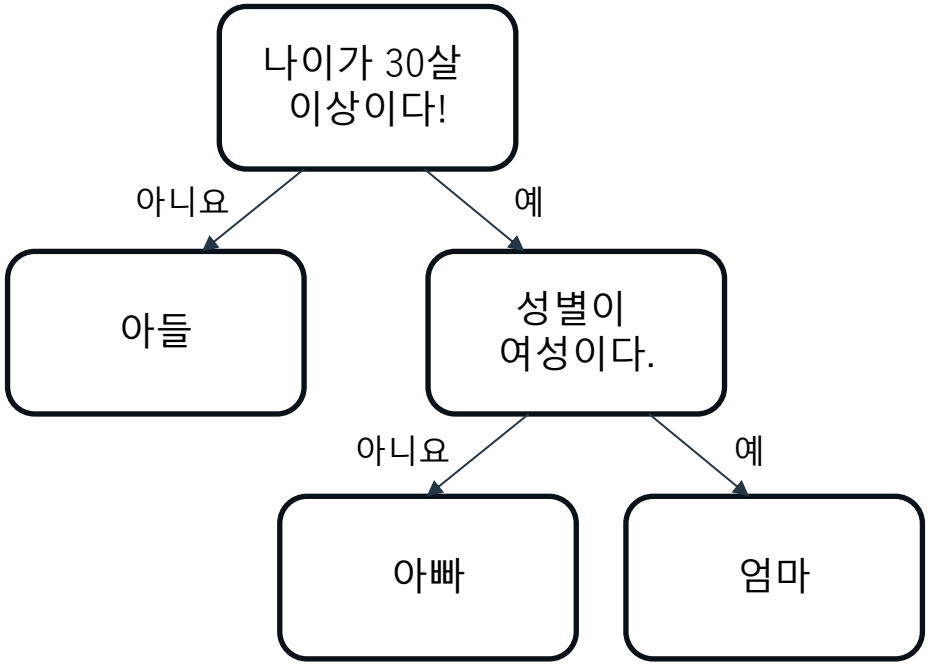
$$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i$$

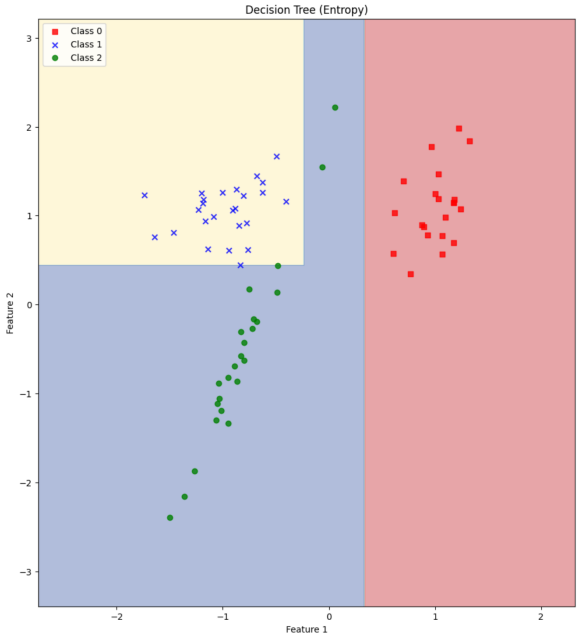
$$subject\ to\ y_i(w^T\phi(x_i) + b) \geq 1 - \xi_i\ and\ \xi_i \geq 0\ for\ 1 \leq i \leq n$$

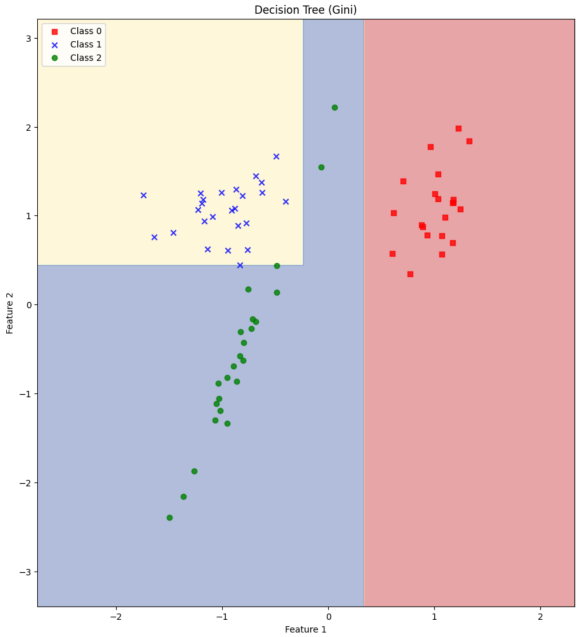


Decision Tree

- 데이터 내에 존재하는 엔트로피 혹은 지니 불순도를 줄이는 방향으로 Tree를 생성
  - $InfoGain = Entropy_{Parent} - \sum \frac{N_{child}}{N_{total}} Entropy_{Child}$
  - $Gini(S) = 1 - \sum_{i=1}^{n} p_i^2$

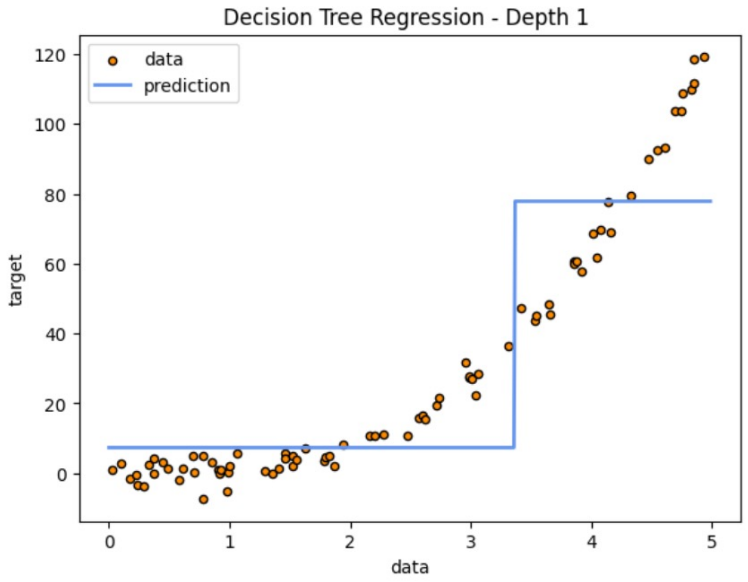


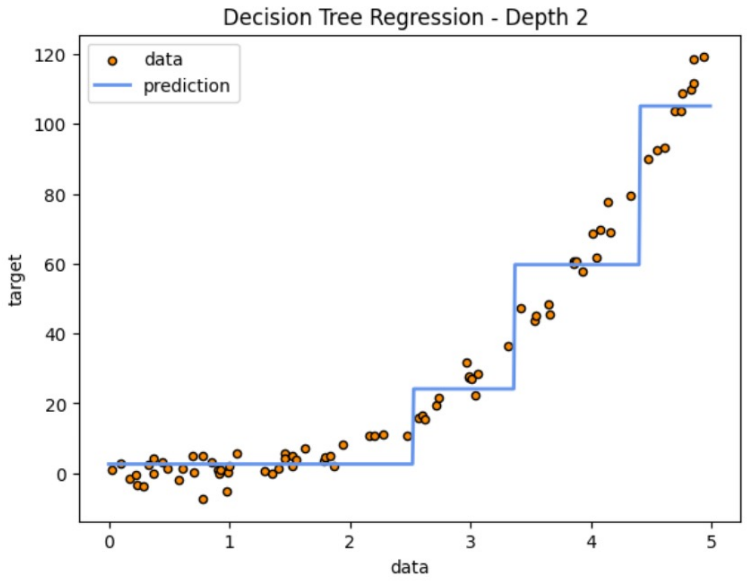


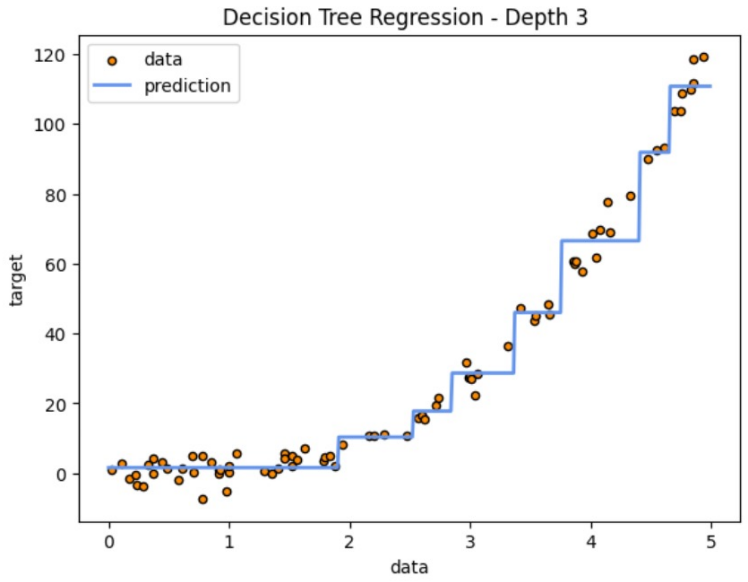


Decision Tree Regression

- 분할 후 실제 정답과 예측 값 사이의 MSE 값이 작아지는 방향으로 노드를 분할







기계가 보도록 만드는 방법 : CNN

- **사람이 보는 과정**
  1. 분석 단위를 설정 후 정보를 추출
  2. 주변 정보를 통합해 차츰 상위 개념을 구성
  3. 목적하는 상위 개념에 도달할 때까지 반복

CNN 이란

- 문제 해결을 위한 특징 (feature) 추출을 Convolution 필터를 사용한 딥러닝 모델
- 또한 Pooling 과정과 반복의 과정으로 상위 개념의 특징을 만드는 과정이 수반

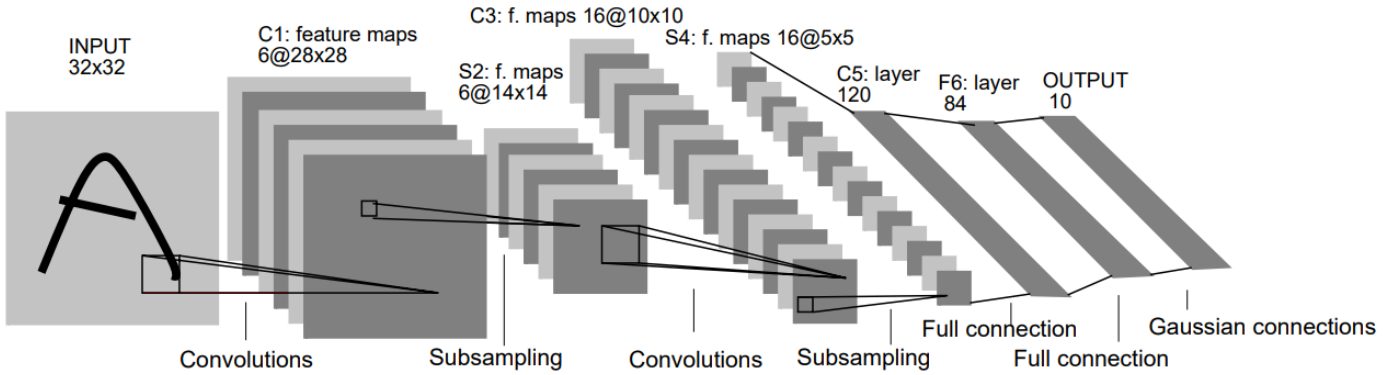


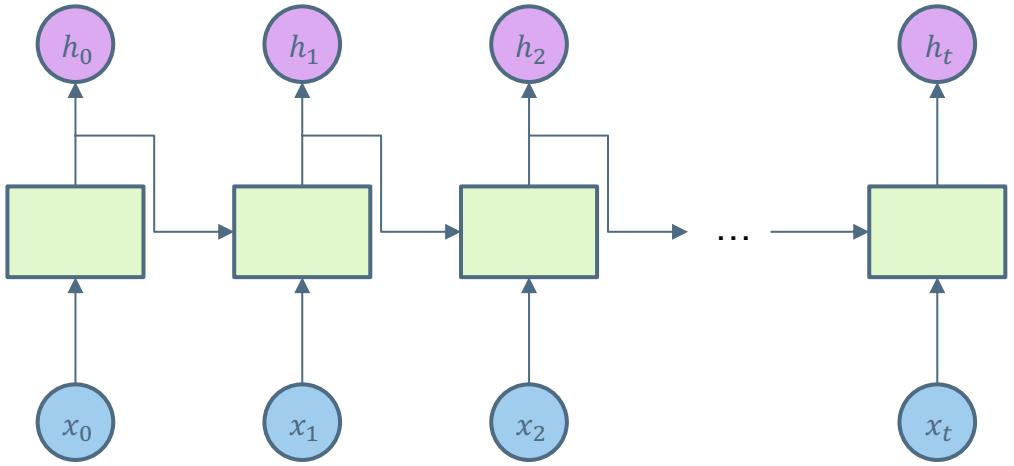
Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

http://vision.stanford.edu/cs598_spring07/papers/Lecun98.pdf

기계도 글을 읽게 만든다 : RNN

- 텍스트와 같은 순차 데이터를 처리하기 위해 고안된 모델
- 이웃한 텍스트 글자 간의 연관성을 표현하는 방식으로 정보를 처리

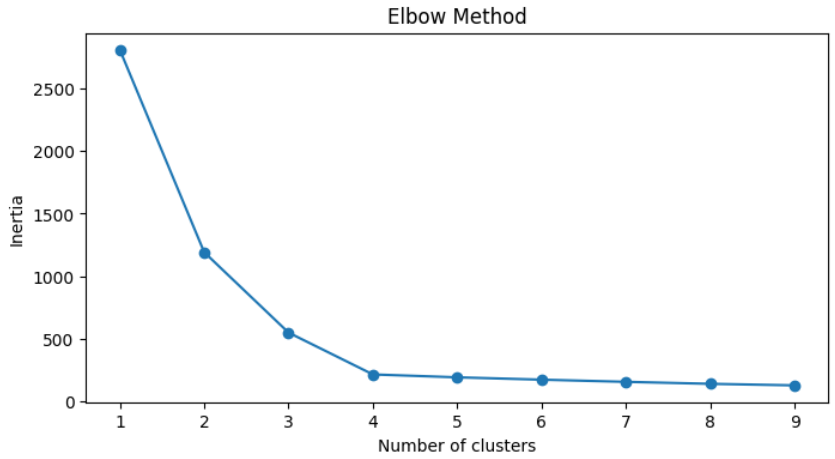
- 사람의 입장의 '기억'을 담당하는 정보 덩어리인 'hidden state'가 존재
- 새로운 입력이 들어오면 이 hidden state를 조금씩 수정



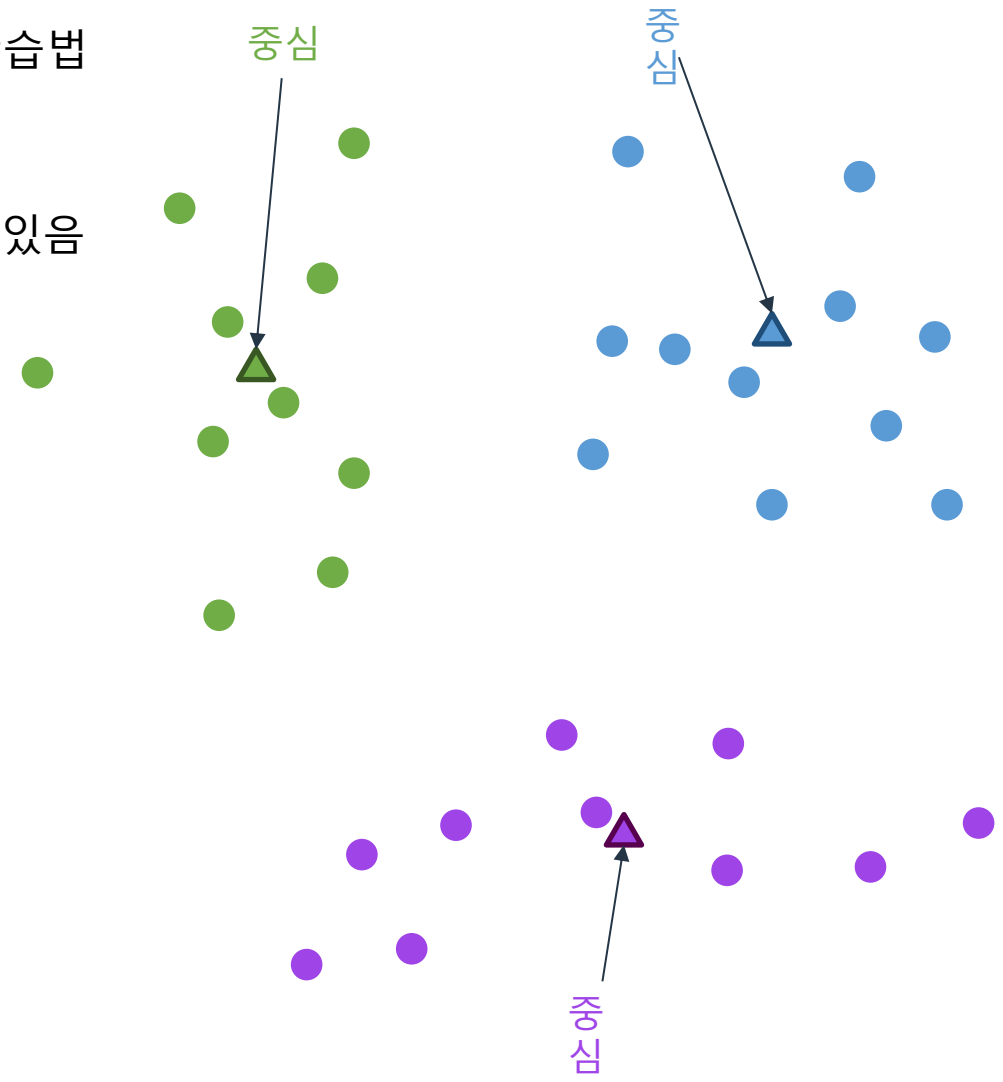
K-means Clustering

- ‘K-평균 군집화’라고 부르며
- 전체 데이터를 K개의 덩어리(클러스터)로 나누는 비지도 학습법

- K 값은 사용자가 정해주는 값으로
- 이를 위한 최적화 방법으로 엘보우 방법(Elbow Method)가 있음

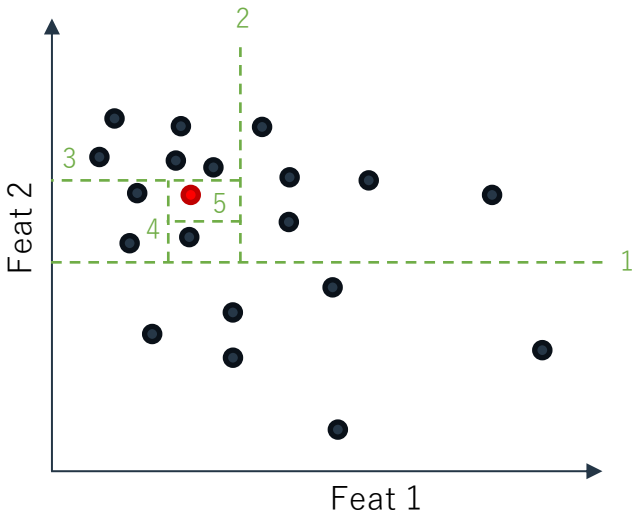


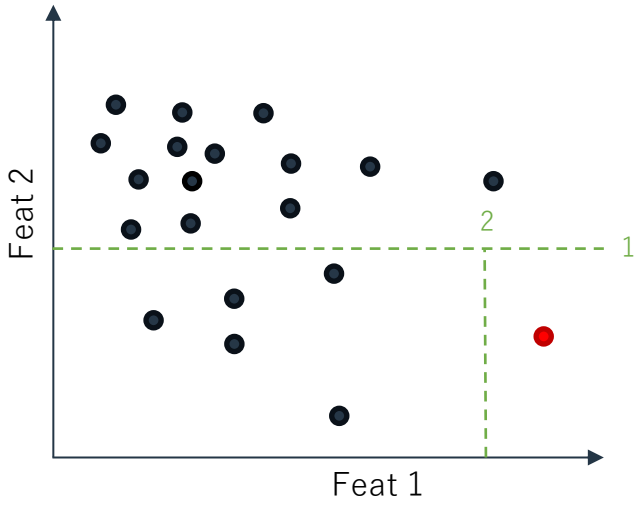
- 평가 metric으로 SSE 혹은 실루엣 계수가 존재



Isolation Forest

- 이때 정상 데이터는
  - 밀도가 높은 구역에 있으니
  - 많은 분할 과정이 지나야 '고립'되며
- 이상치 데이터는
  - 밀도가 낮은 지역에 있어서
  - 낮은 수준의 분할 과정으로도 쉽게 고립시킬 수 있음
- 이러한 데이터의 특성을 이용해 이상치 데이터를 확인하는 방법





E.O.D