

고려대학교
빅데이터 연구회

KU-BIG

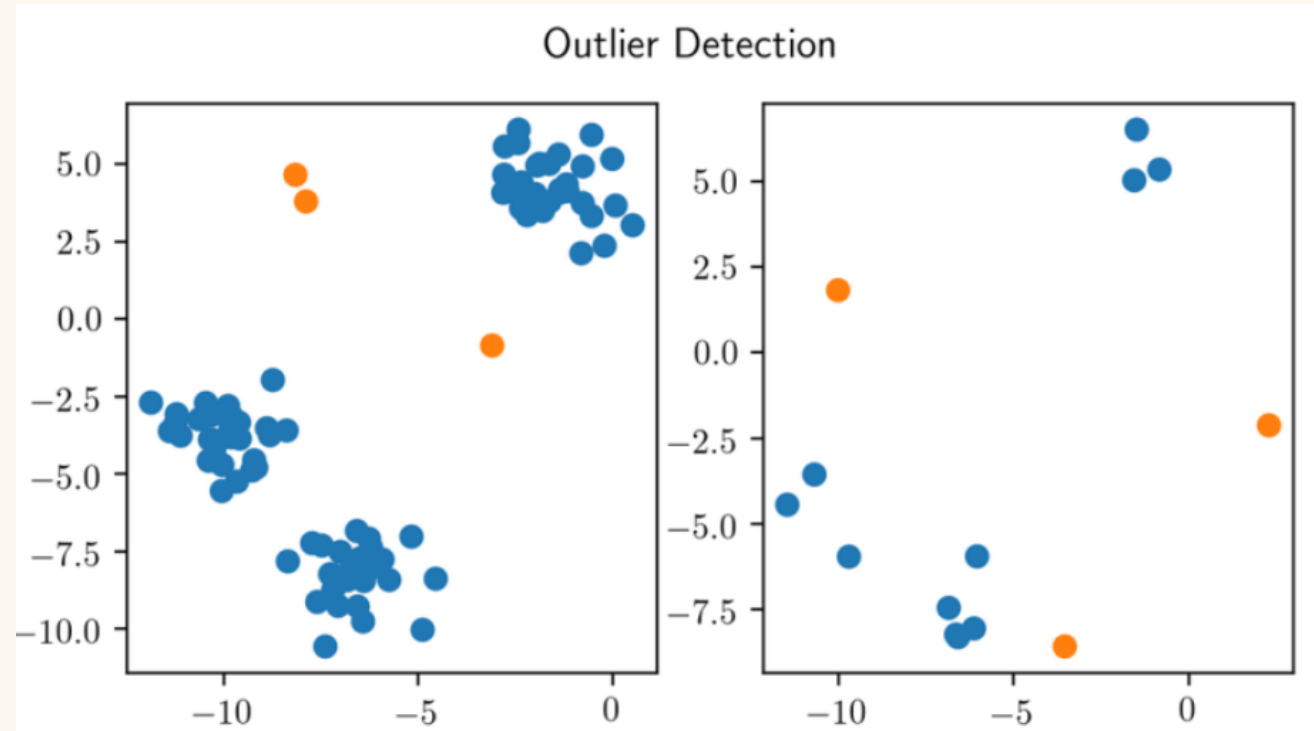
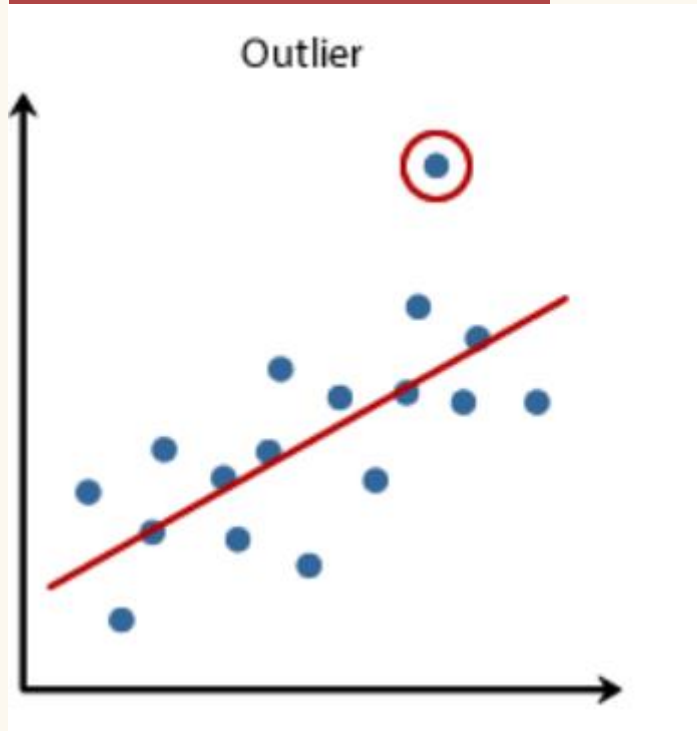
Outlier Detection

유현우 박정진 정희정 송예은 심정은 양수형



Outlier Detection

이상 감지



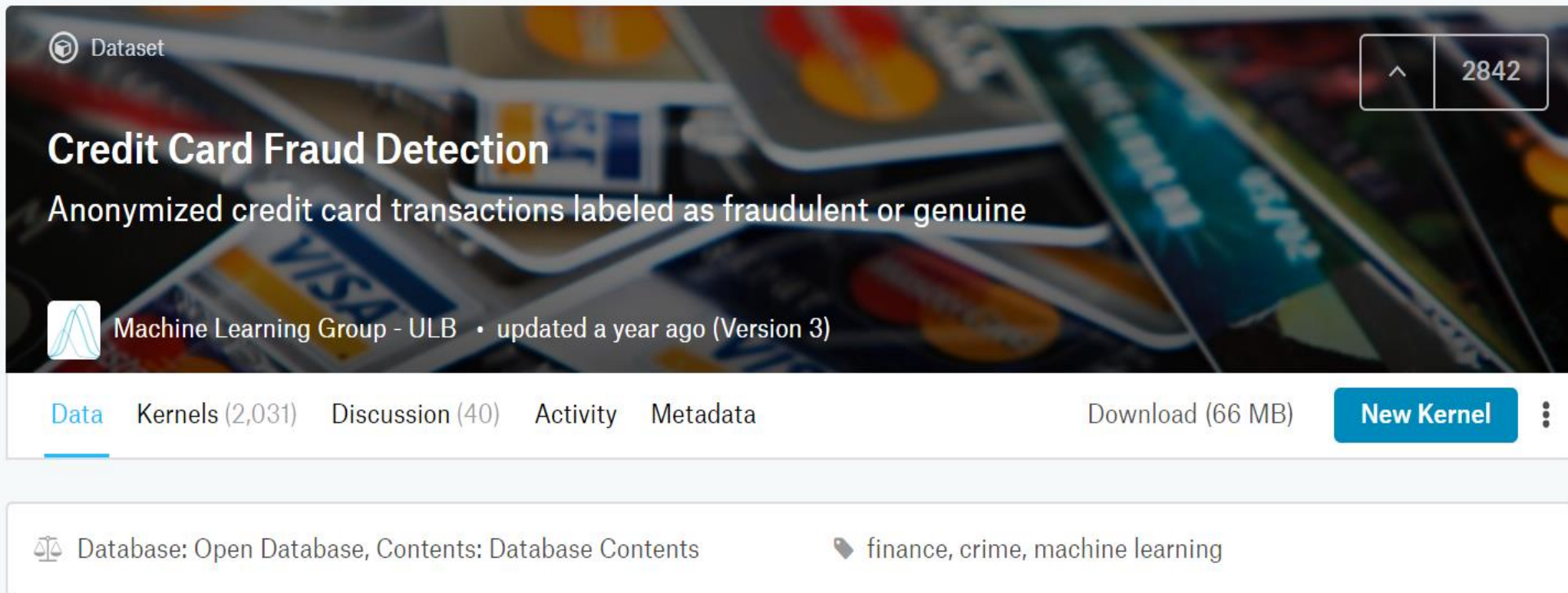
Outlier Detection



COMPUTER SYSTEM
MONITORING



Outlier Detection 활용 데이터




The screenshot shows the Kaggle dataset page for 'Credit Card Fraud Detection'. The background image is a collage of various credit cards, including Visa and Mastercard. The page layout includes a header with the 'Dataset' label and a view count of 2842. The main title 'Credit Card Fraud Detection' is prominently displayed, followed by the description 'Anonymized credit card transactions labeled as fraudulent or genuine'. Below this, the creator 'Machine Learning Group - ULB' is mentioned, along with the update status 'updated a year ago (Version 3)'. A navigation bar contains links for 'Data', 'Kernels (2,031)', 'Discussion (40)', 'Activity', and 'Metadata'. On the right side of this bar, there is a 'Download (66 MB)' link and a 'New Kernel' button. At the bottom, a metadata section shows the database as 'Open Database' and lists tags for 'finance', 'crime', and 'machine learning'.

Dataset



2842

Credit Card Fraud Detection

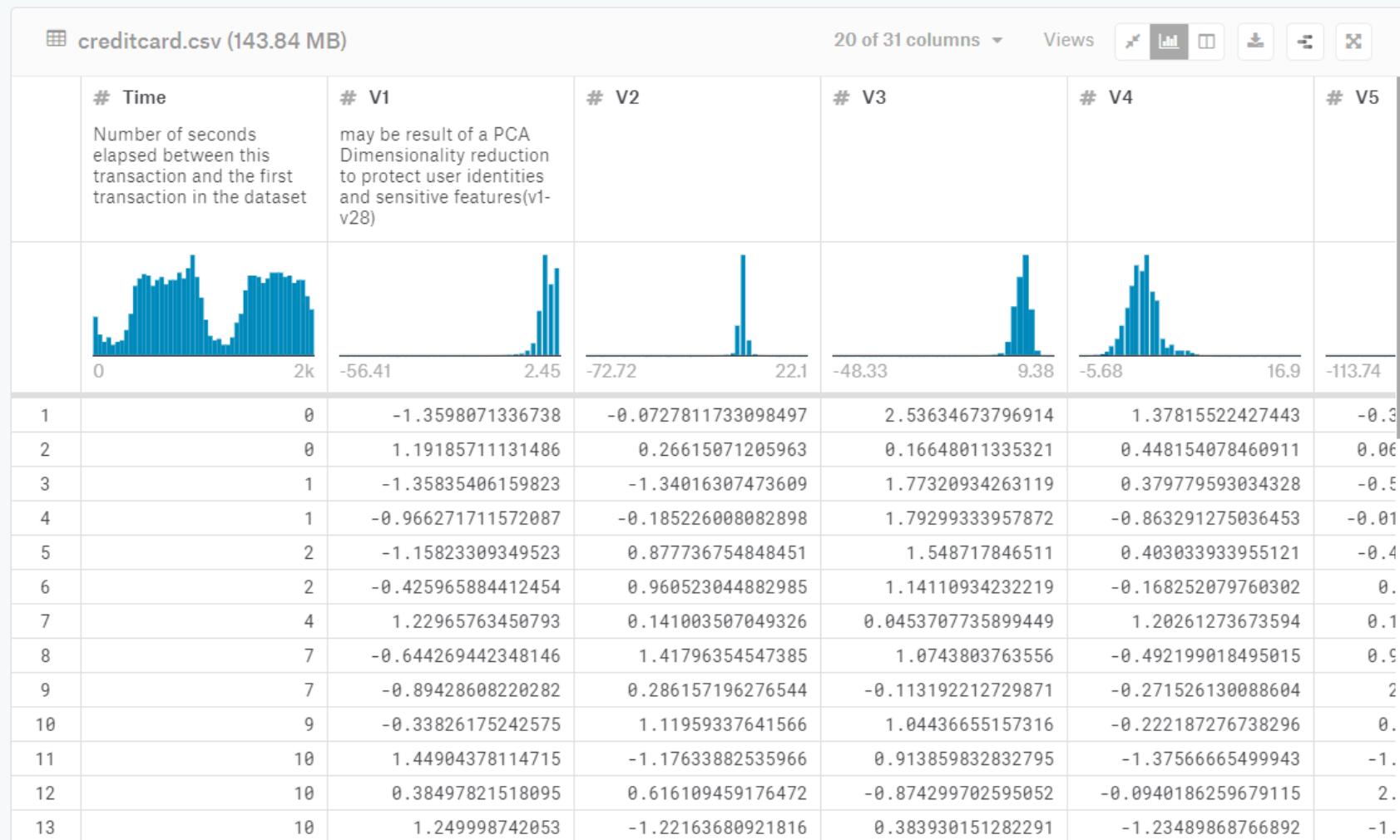
Anonymized credit card transactions labeled as fraudulent or genuine

 Machine Learning Group - ULB • updated a year ago (Version 3)

[Data](#) [Kernels \(2,031\)](#) [Discussion \(40\)](#) [Activity](#) [Metadata](#) [Download \(66 MB\)](#) [New Kernel](#)

 Database: Open Database, Contents: Database Contents  finance, crime, machine learning

Dataset Variables



Columns

- # Time Number of seconds elapsed between
- # V1 may be result of a PCA Dimensionality r
- # V2
- # V3
- # V4
- # V5
- # V6
- # V7
- # V8
- # V9
- # V10
- # V11
- # V12
- # V13
- # V14
- # V15
- # V16
- # V17
- # V18
- # V19
- # V20
- # V21
- # V22
- # V23
- # V24
- # V25
- # V26
- # V27
- # V28 abc
- # Amount Transaction amount
- ✓ Class 1 for fraudulent transactions, 0 other

Dataset Variables

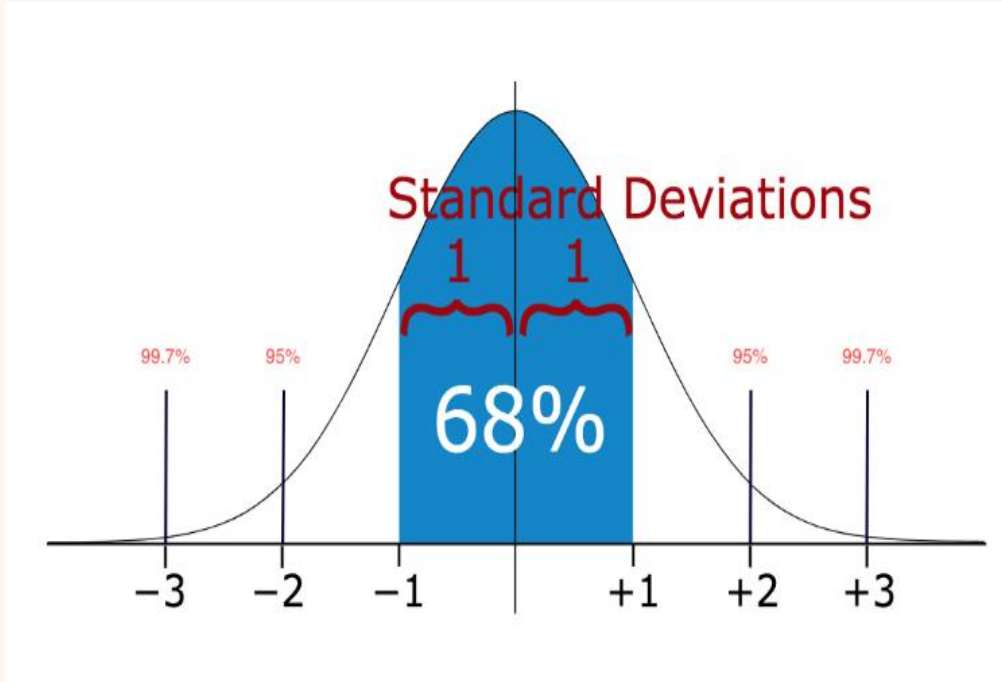
Time Number of seconds elapsed between this transaction and the first transaction in the dataset

Amount Transaction amount

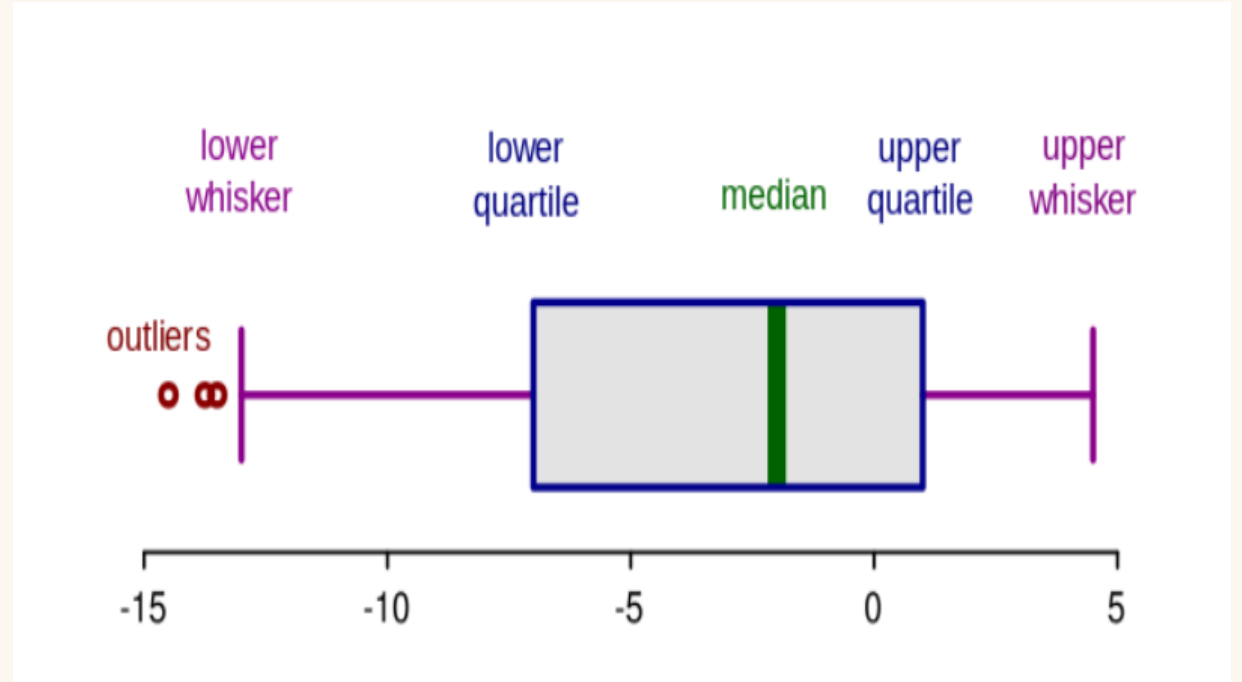
V1 may be result of a PCA Dimensionality reduction to protect user identities and sensitive features(v1-v28)

✓ Class 1 for fraudulent transactions, 0 otherwise

Statistical Method



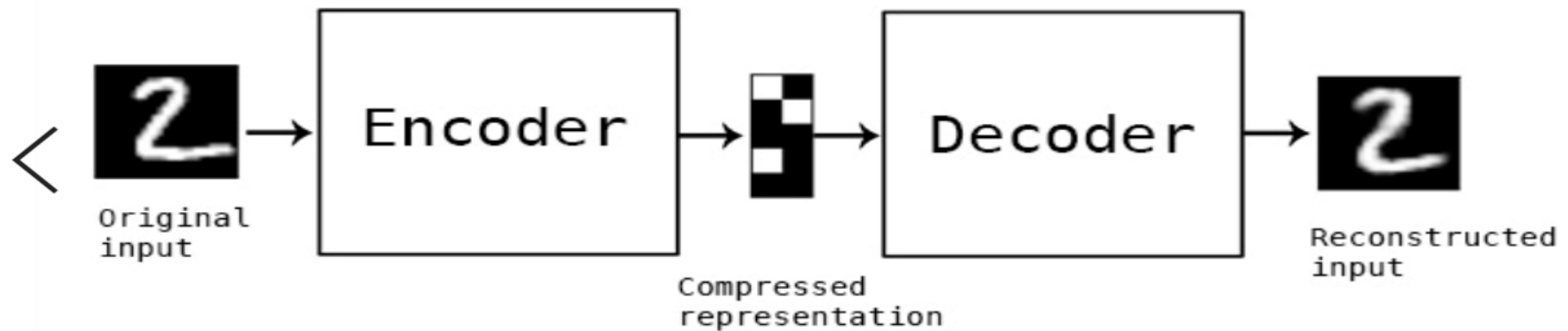
Standard Deviations



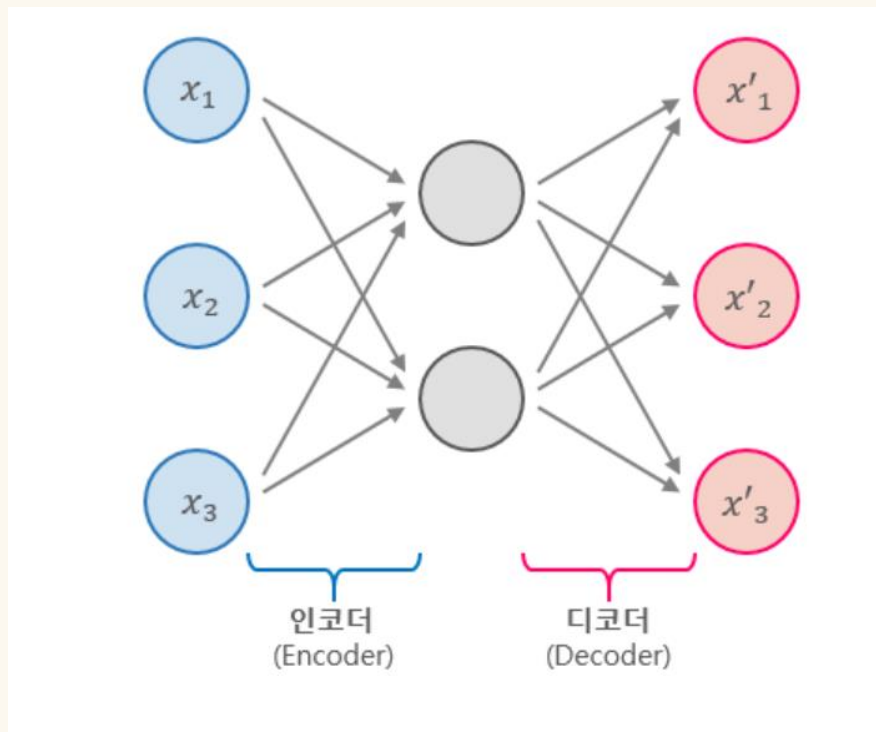
Boxplots

Autoencoder

입력 → 압축 → 압축된 데이터로부터 복원하는 신경망



Autoencoder 방법들



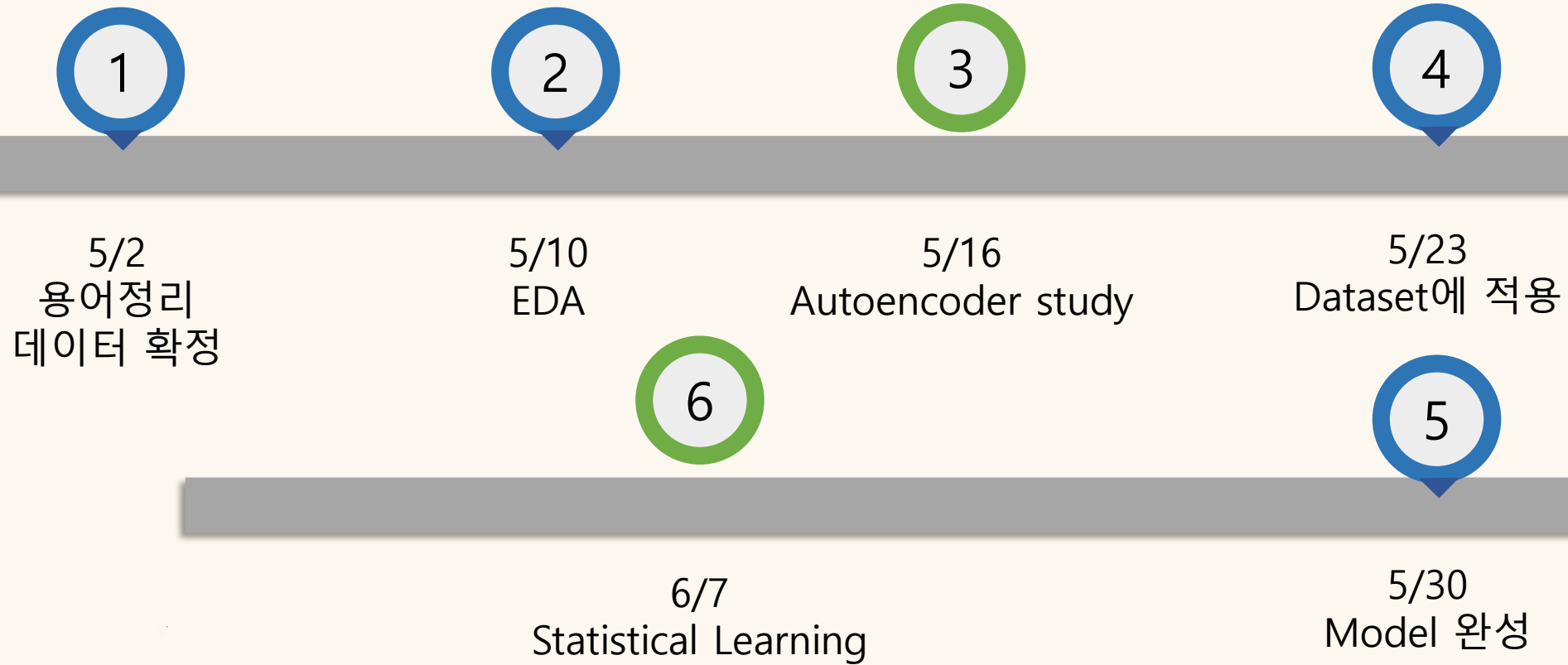
- Hidden layer 수 < input layer 수
→ 데이터 압축
- 입력 데이터에 noise 추가
→ 입력을 복원 가능하도록 학습
- VAE, Stacked Autoencoder 등

Autoencoder

Autoencoder 제약의 효과?

단순히 입력을 바로 출력으로 복사하지 못하도록 방지
데이터를 효율적으로 표현하는 방법을 학습하도록 제어

진행 계획



A stylized illustration of a person from the chest up, wearing a grey suit jacket, a white shirt, and a dark tie. The person's face is partially visible at the top, showing a red nose and a brown beard. A large, black-outlined speech bubble is positioned in front of the person's chest. Inside the speech bubble, the text "Do you have any question?" is written in a sans-serif font. The word "question?" is in a larger, pink font, while "Do you have any" is in a smaller, grey font.

Do you
have any
question?

Thank you
for your attention.