

Lecture #17: Classification Using Scikit Learn

- loading datasets (Titanic Dataset)
- classification (classes identify)
 - categorical
 - binary class
 - multi-class

- importing libraries

Sklearn for me much easier
matplotlib, seaborn ⇒ for visualization
numpy, pandas, ⇒ data pre-processing

- revision of seaborn tutorial

① loading data (`pd.read_csv('')`)

② understanding data info

- shape
- `head()`

③ understanding through Charts

(to get better insights) • Distribution plot

distplot

histplot

• Relational plot

relplot

scatterplot

(effective, intelligent plotting)

• Categorical Plot

box plot

(tell us five number summary)

violin plot

bar plot

strip plot

swarm plot

Q1

Q2

Q3

min, max

Q3 - Q1

④ Data cleaning

null values

fill missing value (imputation)

drop columns, rows

obied to convert & categorize data types

⇒ Getting model ready [all the data is numerical
everything is concatenated together

⇒ Machine learning

↳ split the dataset into input & target variable.
(features) ↓ (Target)