Applications

Of Data Science

Topics in Classification

Applications of Data Science - Class 13

Giora Simchoni

`gsimchoni@gmail.com and add #dsapps in subject`

Stat. and OR Department, TAU

2020-03-01

Topics in Classification

Life isn't perfect

Let's tackle just a few issues:

- Imbalanced Classes
- Multiple Classes
- Not enough labelled data and data labeling is expensive

Imbalanced Classes

Typical examples of Imbalanced Classes scenarios

- Rare diseases: [this](#) dataset contains genetic data for 1,000 HIV patients, 206 out of 1,000 patients improved after 16 weeks of therapy

- Conversion/Sell/CTR rates: [this](#) dataset contains 10 days of Click-Through-Rate data for Avazu mobile ads, ~6.8M clicked out of ~40.4M

- Fraud detection: [this](#) dataset contains credit card transactions for a major European CC, 492 frauds out of 284,807 transactions