



UNIVERSIDADE
CATÓLICA
PORTUGUESA

BRAGA

Machine Learning

Session 22 - PL

Data Imbalance in Machine Learning

Degree in Applied Data Science

2024/2025

Balance Data with imbalanced-learn

- Imbalanced-learn (imported as imblearn) is an open source, MIT-licensed library relying on scikit-learn (imported as sklearn) and provides tools when dealing with classification with imbalanced classes.
- Installation:
 - `pip install imbalanced-learn`OR
 - `conda install -c conda-forge imbalanced-learn`

Balance Data with imbalanced-learn

1.1. API's of imbalanced-learn samplers

The available samplers follows the scikit-learn API using the base estimator and adding a sampling functionality through the `sample` method:

Estimator: The base object, implements a `fit` method to learn from data, either:

```
estimator = obj.fit(data, targets)
```

Resampler: To resample a data sets, each sampler implements:

```
data_resampled, targets_resampled = obj.fit_resample(data, targets) 
```

Exercises:

- Notebooks on the github repository:
 - Notebook with examples and exercises:
 - `exercises/session22/examples.ipynb`
 - `exercises/session22/exercises.ipynb`