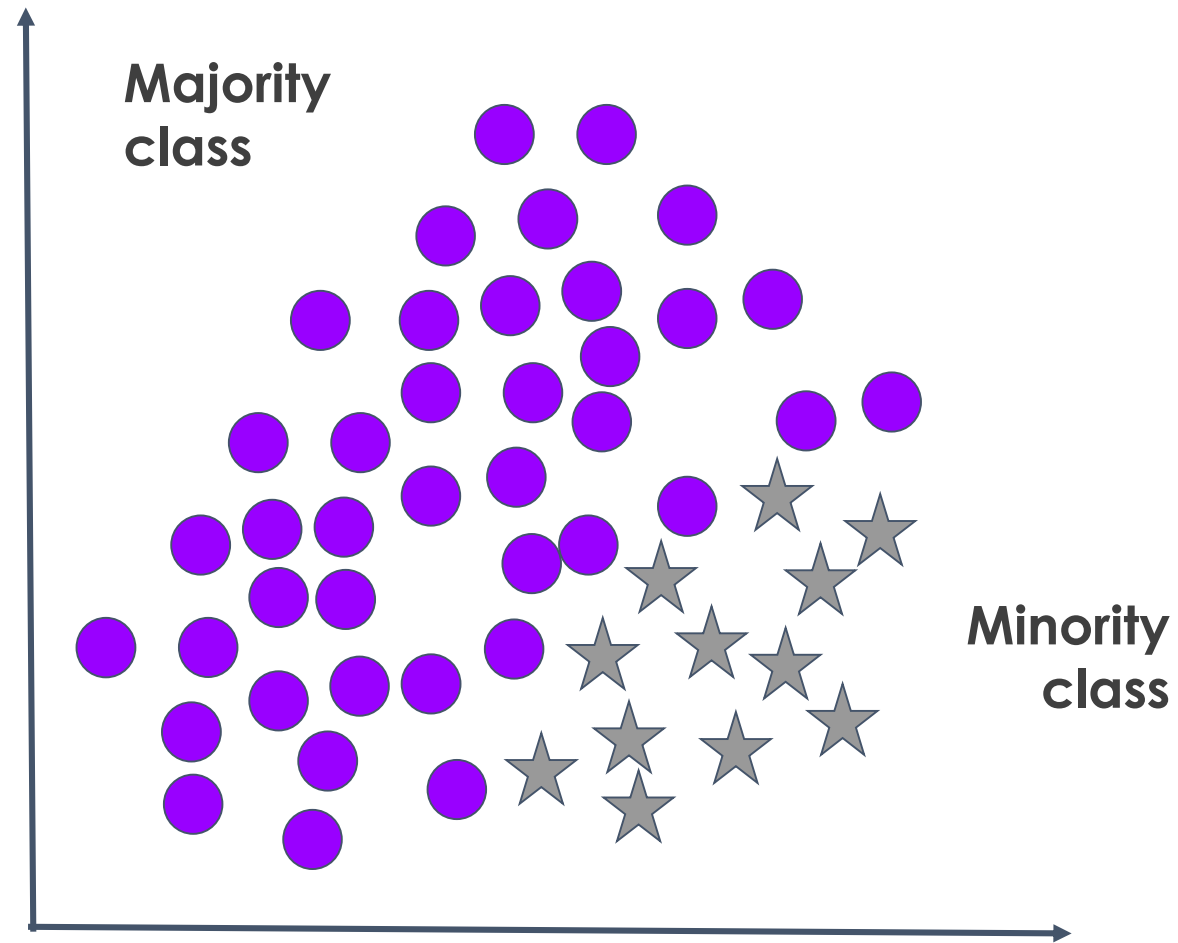# Tomek Links

Tomek Links

If 2 samples are nearest neighbours, and from a different class, they are Tomek Links.

Removes the Tomek Link from the majority class
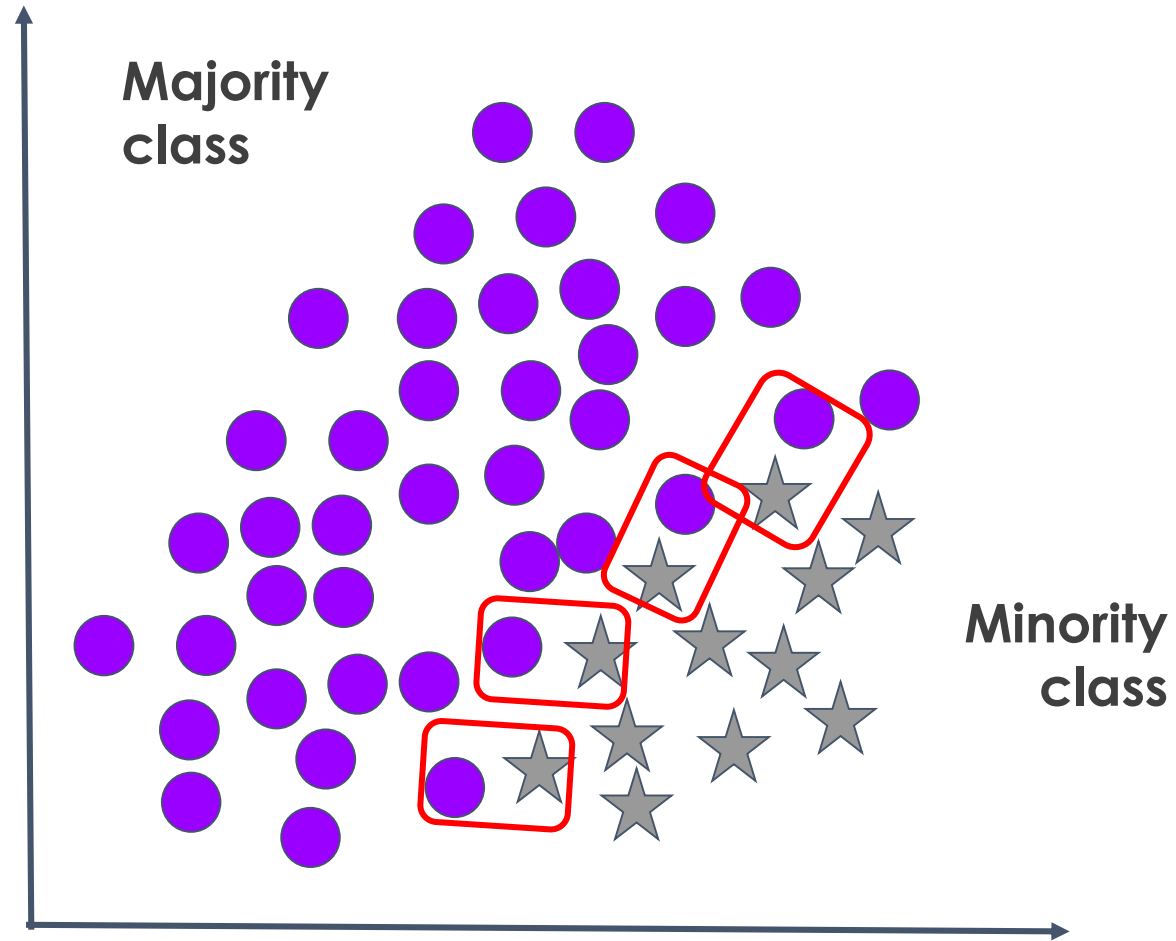
- Cleaning

- Final dataset shape varies

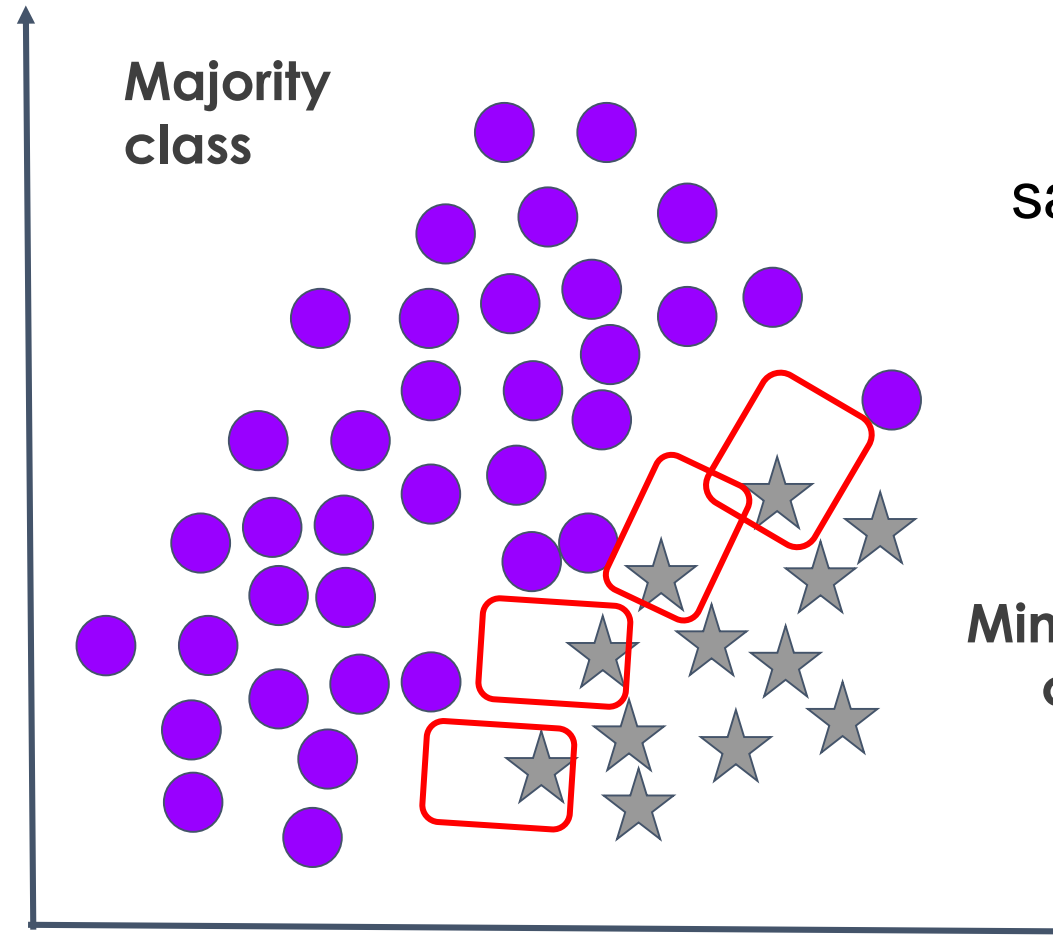- Boundary is noise

Tomek Links
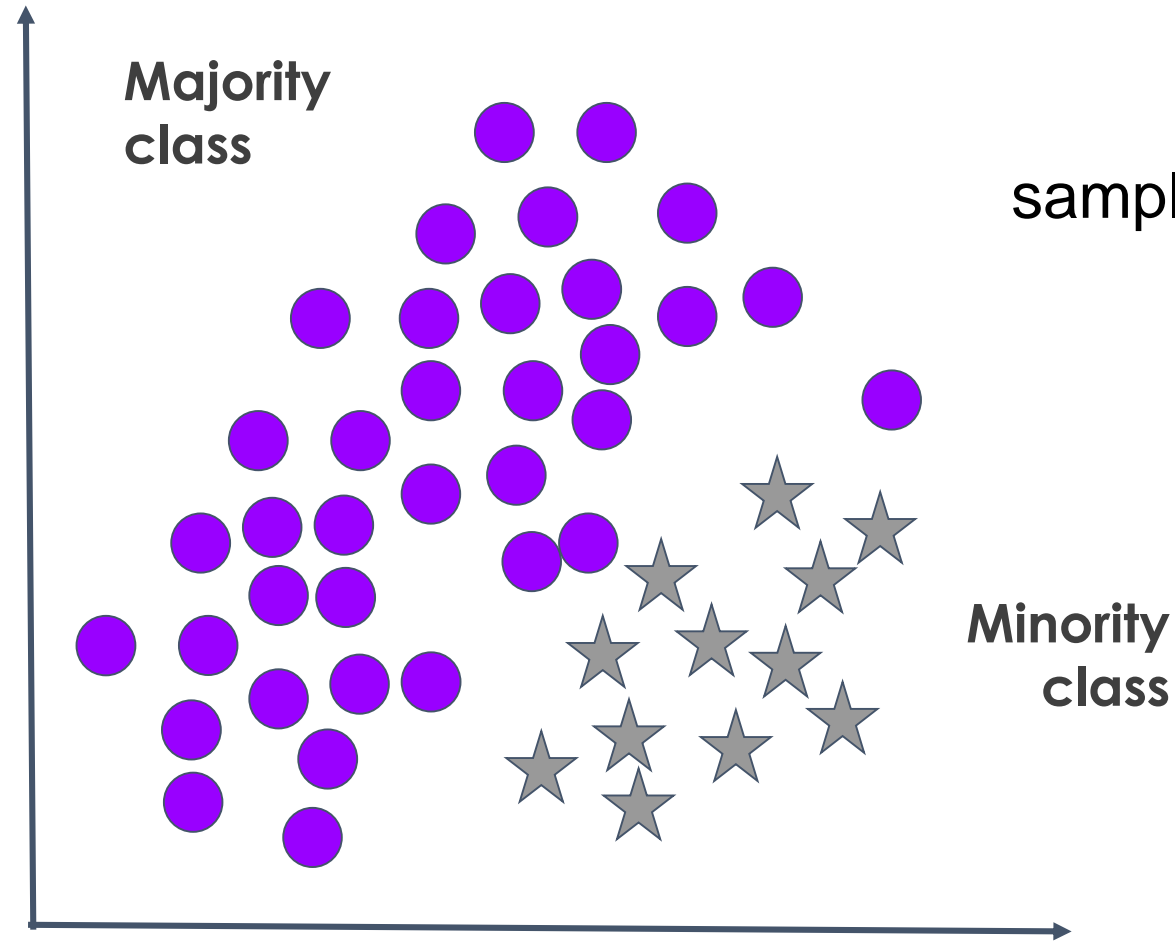
Tomek Links

Find Tomek Links

Tomek Links

Remove observation from majority class in Tomek Link

Majority class
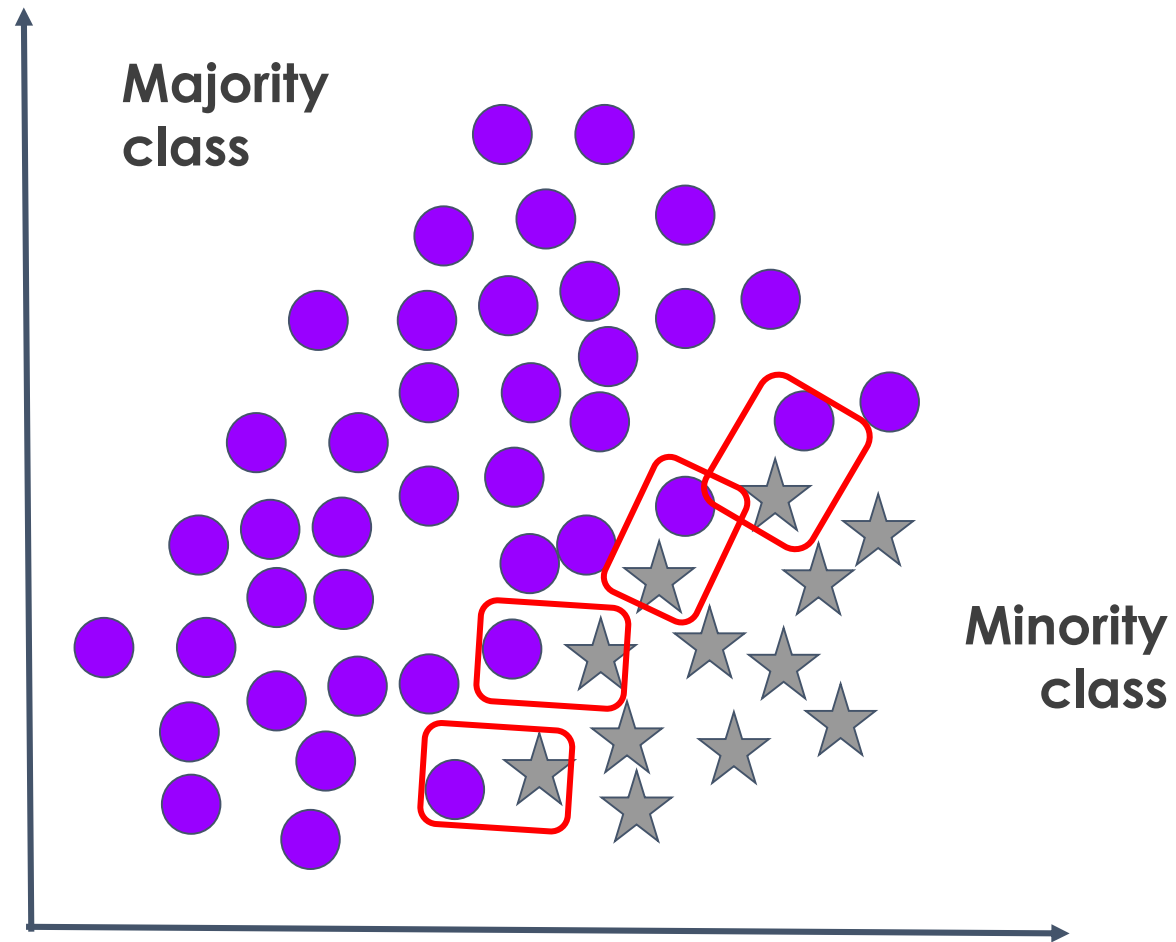
Minority class

sample_strategy = 'auto'

Tomek Links

Remove observation from majority class in Tomek Link

sample_strategy = 'auto'

Majority class

Minority class

Tomek Links

Find Tomek Links

Majority class

Minority class

Tomek Links

Remove entire
Tomek Link

sample_strategy = 'all'

Majority class

Minority class

Tomek Links

Remove entire
Tomek Link

Majority
class
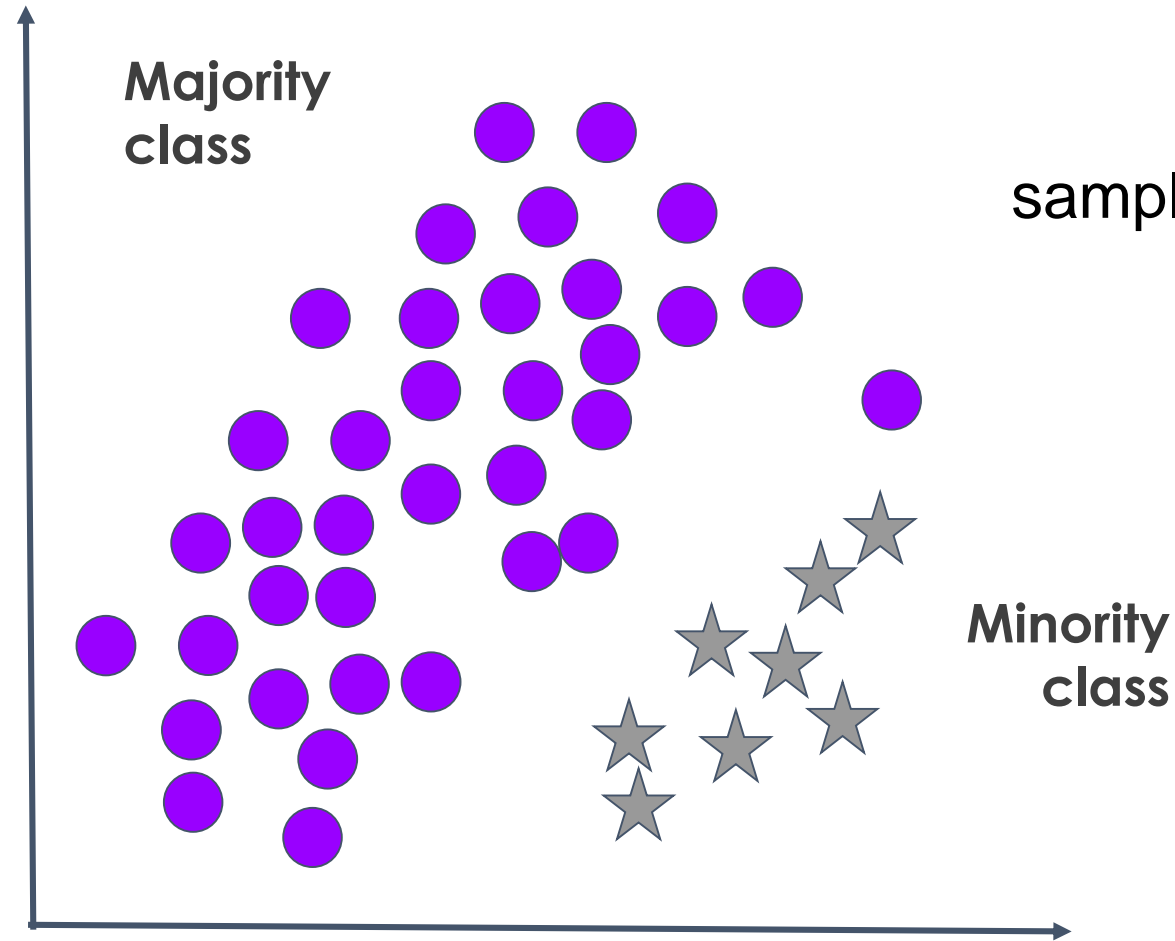
sample_strategy = 'all'

Minority
class

Condensed Nearest Neighbours (CNN)

Removes noise ➜ improves performance

Miss classifies hard cases
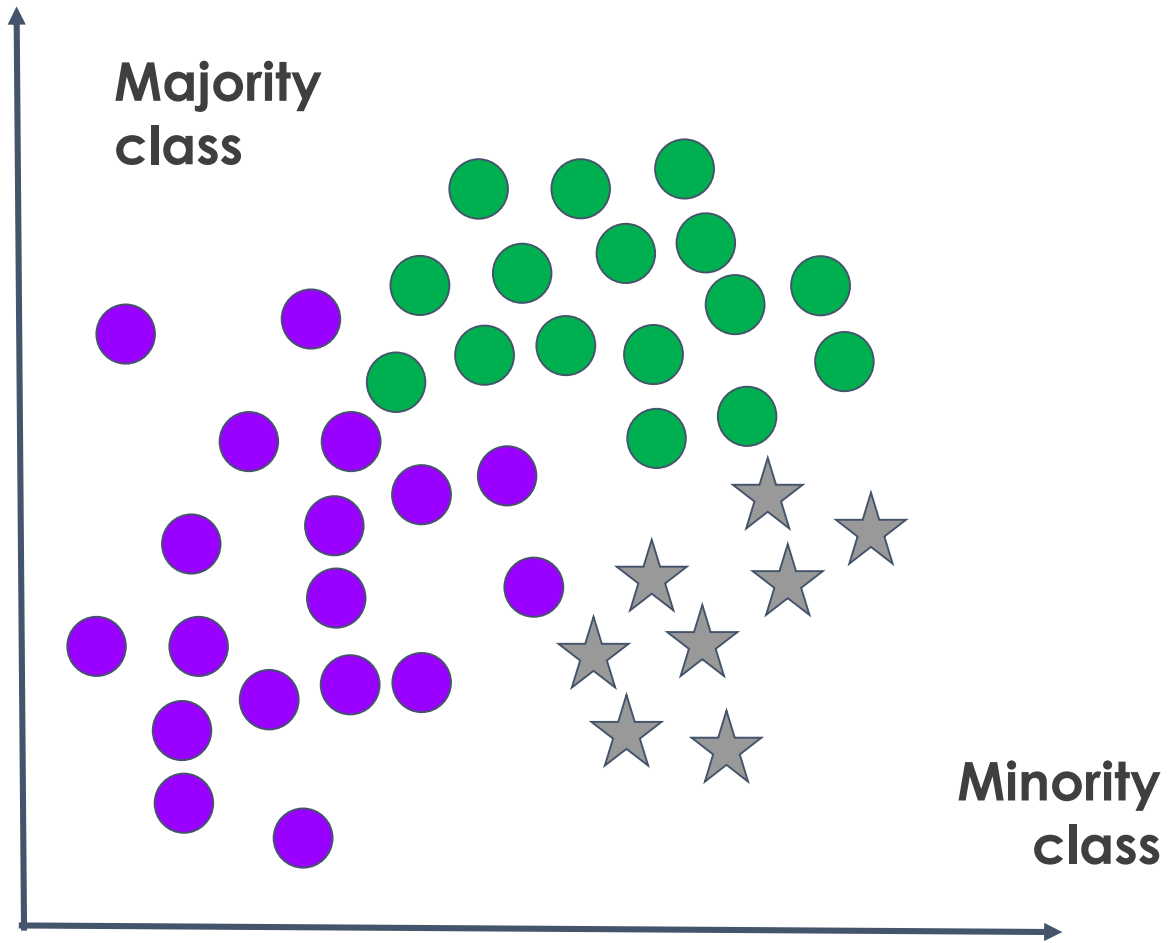
Imbalanced-learn: Tomek Links

```python
# create data
X, y = make_data(sep=0.5)

# set up Tomek Links

tl = TomekLinks(
    sampling_strategy='auto',  # undersamples only the majority class
    n_jobs=4)  # I have 4 cores in my laptop

X_resampled, y_resampled = tl.fit_resample(X, y)
```
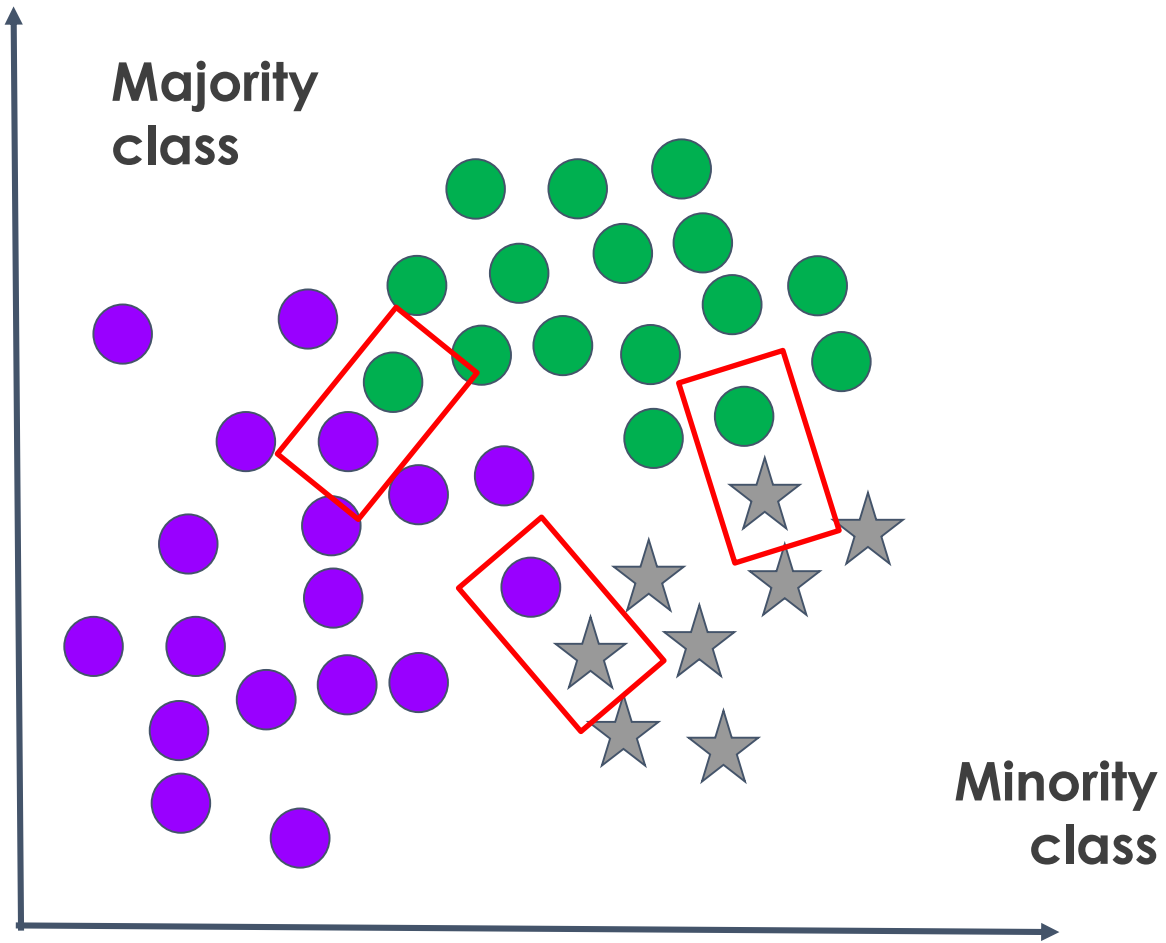
Multi-class

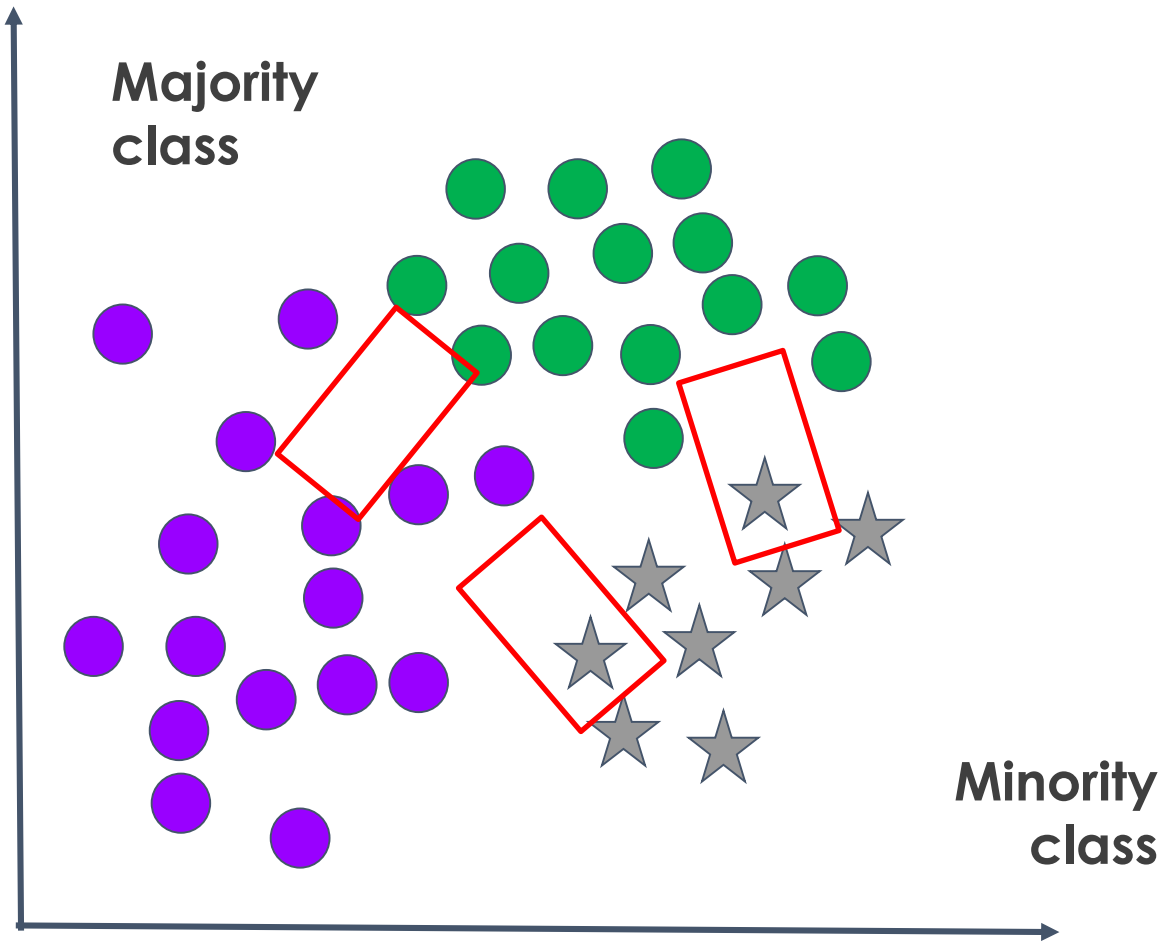Majority class

Minority class

One vs Rest

Multi-class

One vs Rest

Find Tomek Link
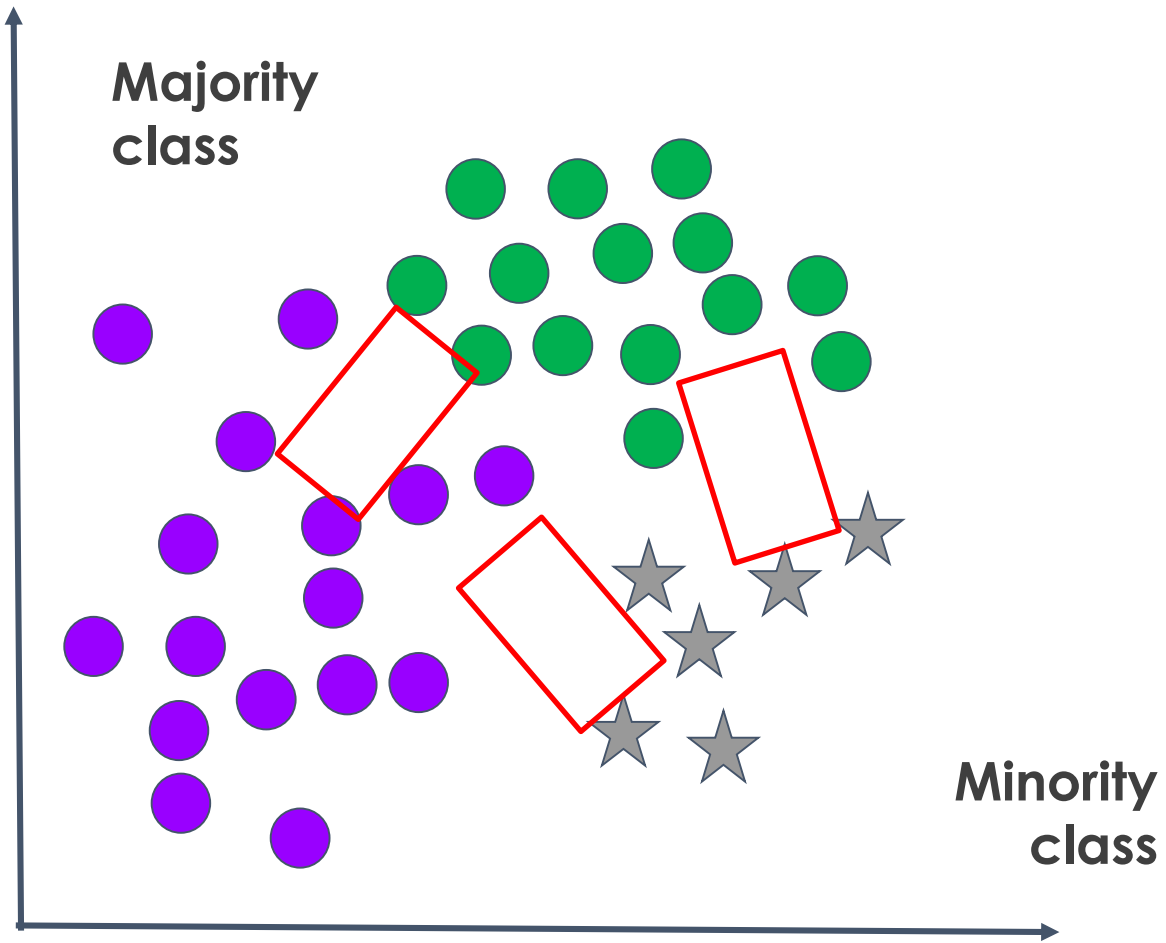
Multi-class

Majority class

Minority class

One vs Rest

Remove sample from majority.

Multi-class

One vs Rest

Remove entire Link.

THANK YOU

www.trainindata.com