



Edited Nearest Neighbours



Edited Nearest Neighbours

Remove samples from the majority class that are closest to the boundary (with the other classes).

Enhance the separation of the classes.



Edited Nearest Neighbours

Removes observations whose neighbours disagree with it on the class.

Typically, 3 neighbours per observation are evaluated.

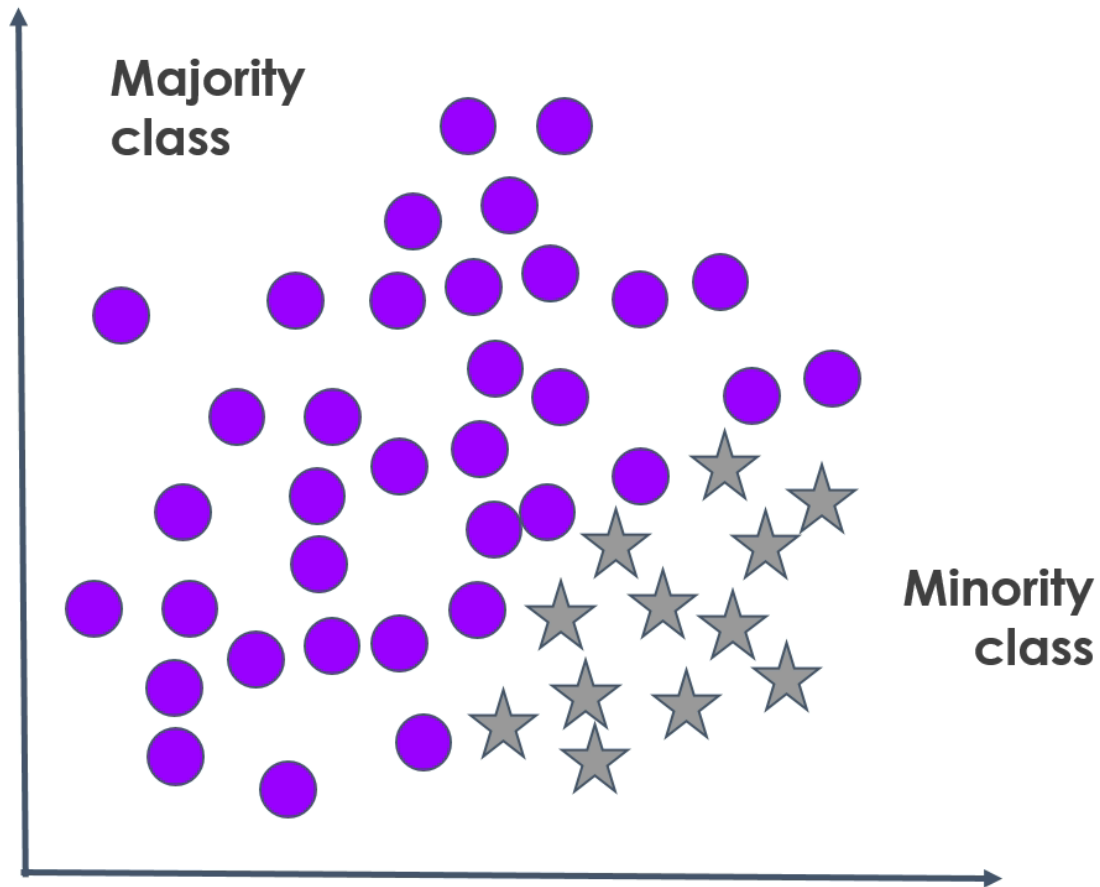
Edited Nearest Neighbours

- Final dataset shape varies
- Cleaning method
- Removes hard cases

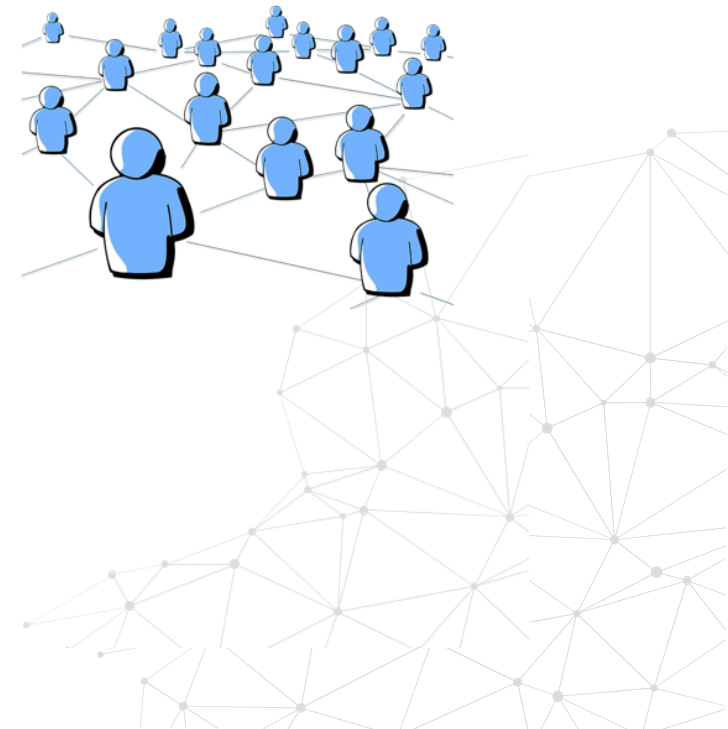
Edited Nearest Neighbours

- 1) Trains a 3 KNN on entire dataset.
- 2) Finds each observation's 3 closest neighbours.
- 3) Keeps or removes observation based on neighbours agreement with its class.
 - Two selection criteria:
 - **All** neighbours need to agree to retain observation
 - **Most** neighbours need to agree to retain observation

Edited Nearest Neighbours

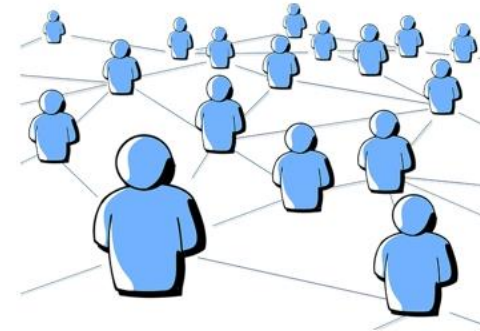
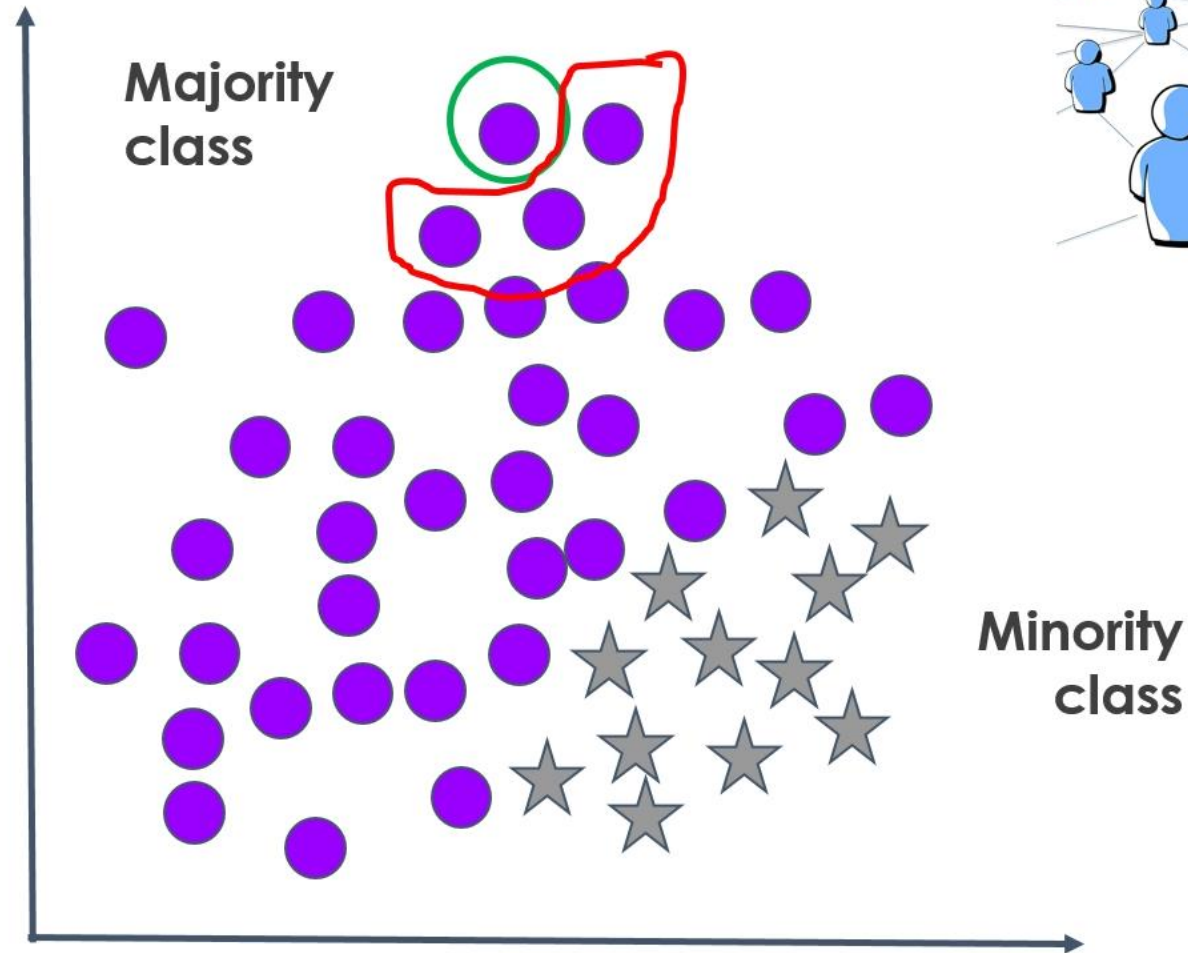


Train a 3 KNN algorithm



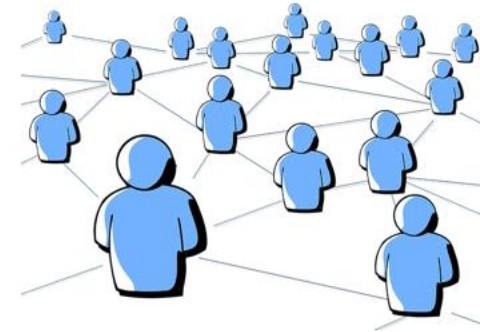
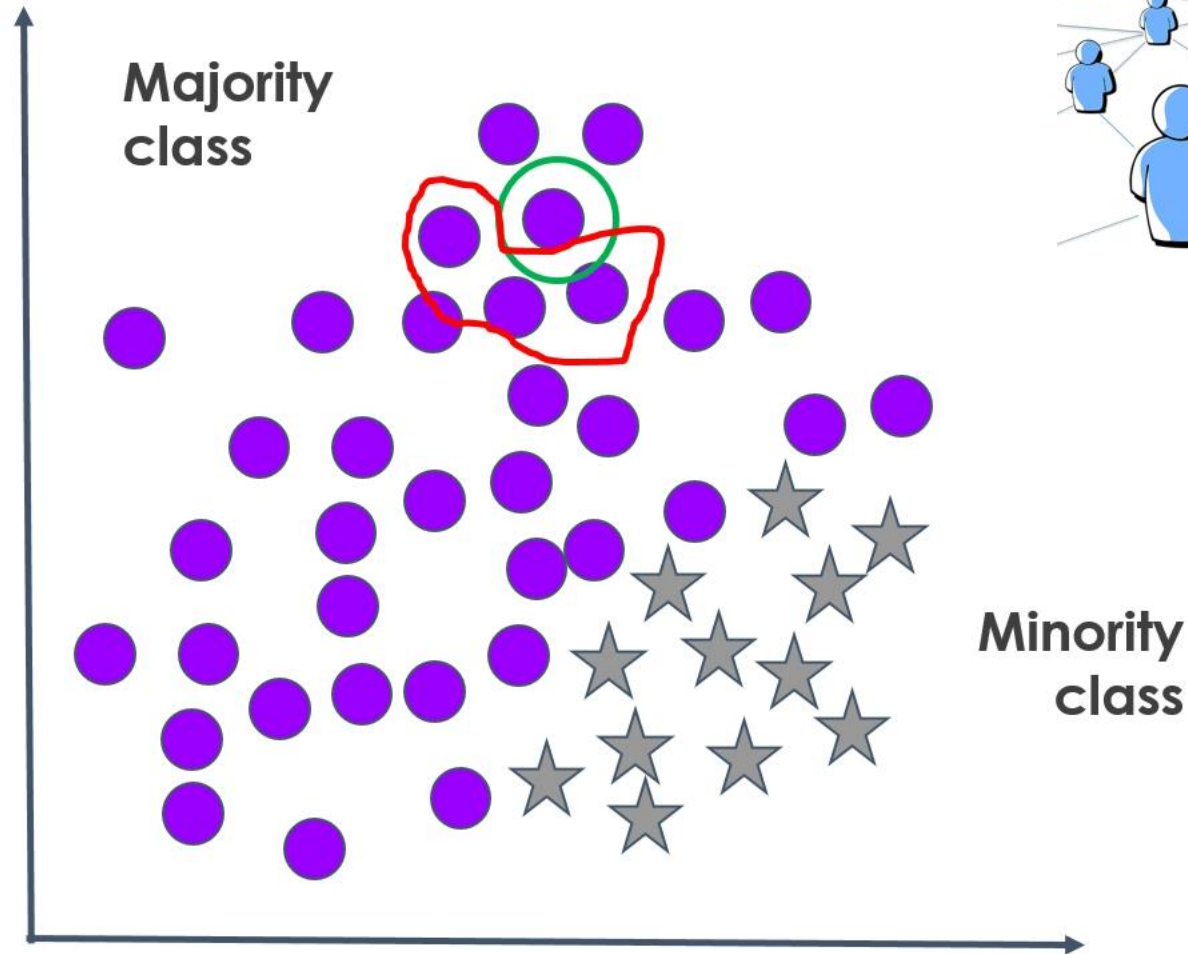
Edited Nearest Neighbours

If neighbours show same class, the observation is kept in the dataset.



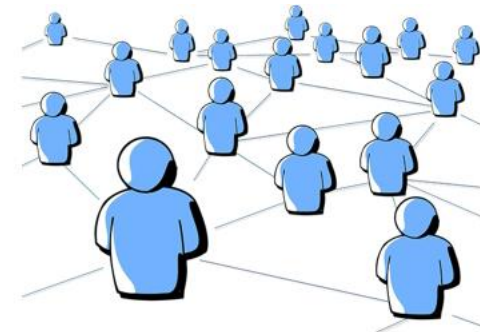
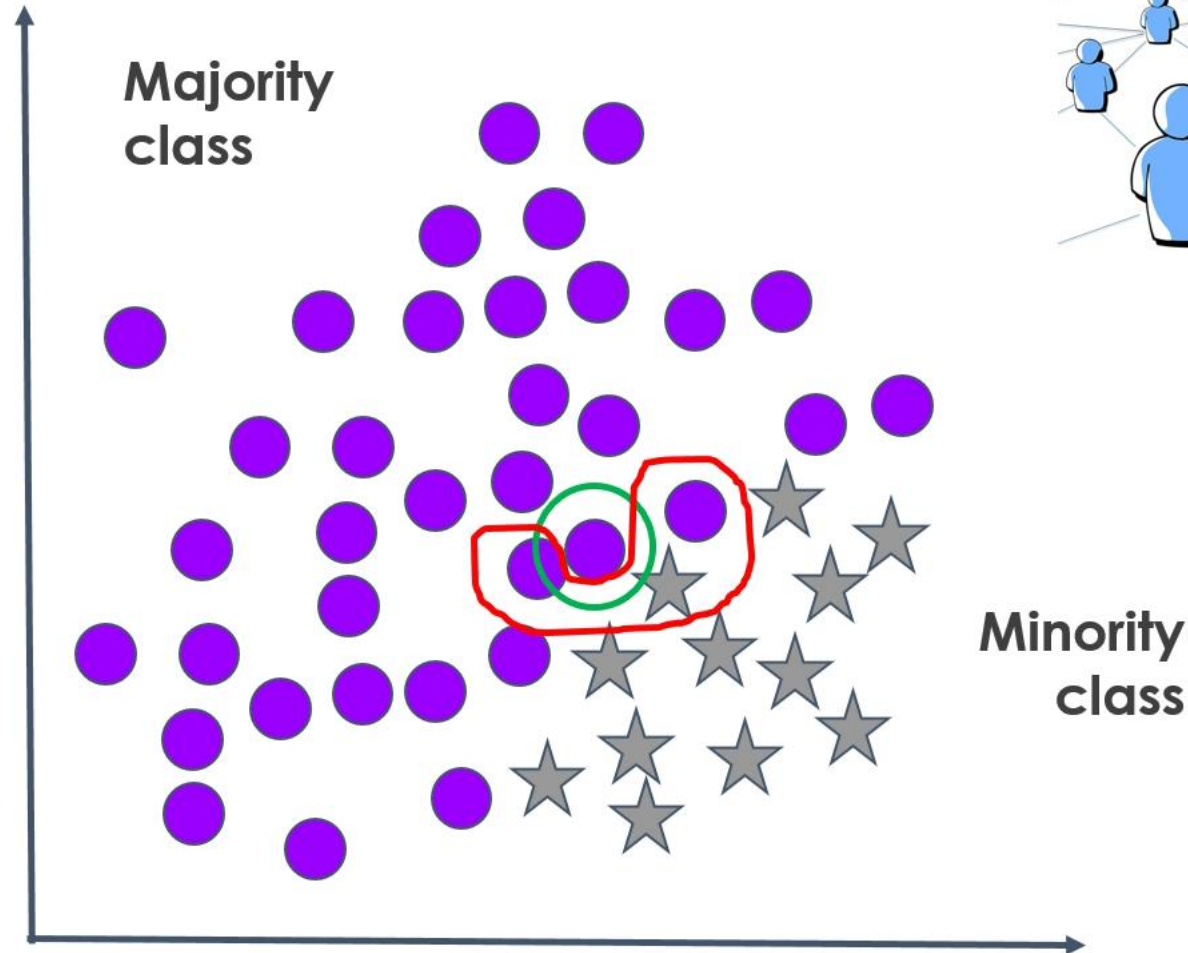
Edited Nearest Neighbours

If neighbours show same class, the observation is kept in the dataset.



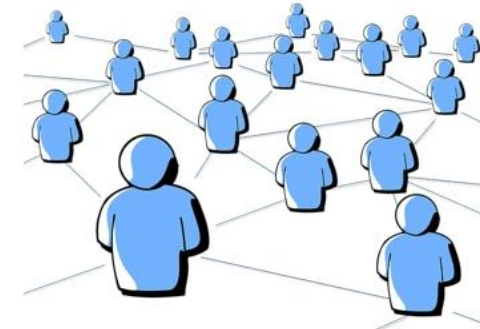
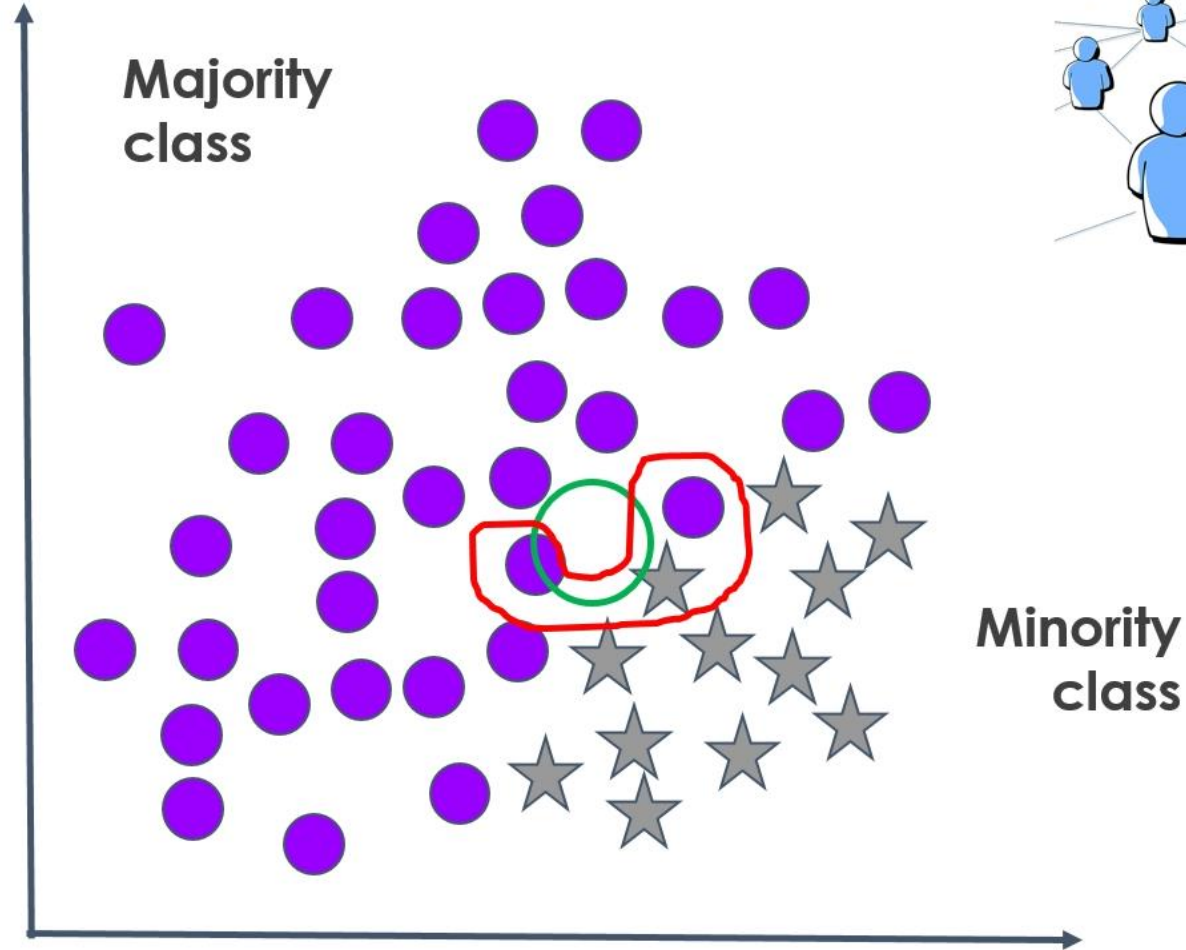
Edited Nearest Neighbours

If some neighbours show a different class, the observation is removed from the dataset.



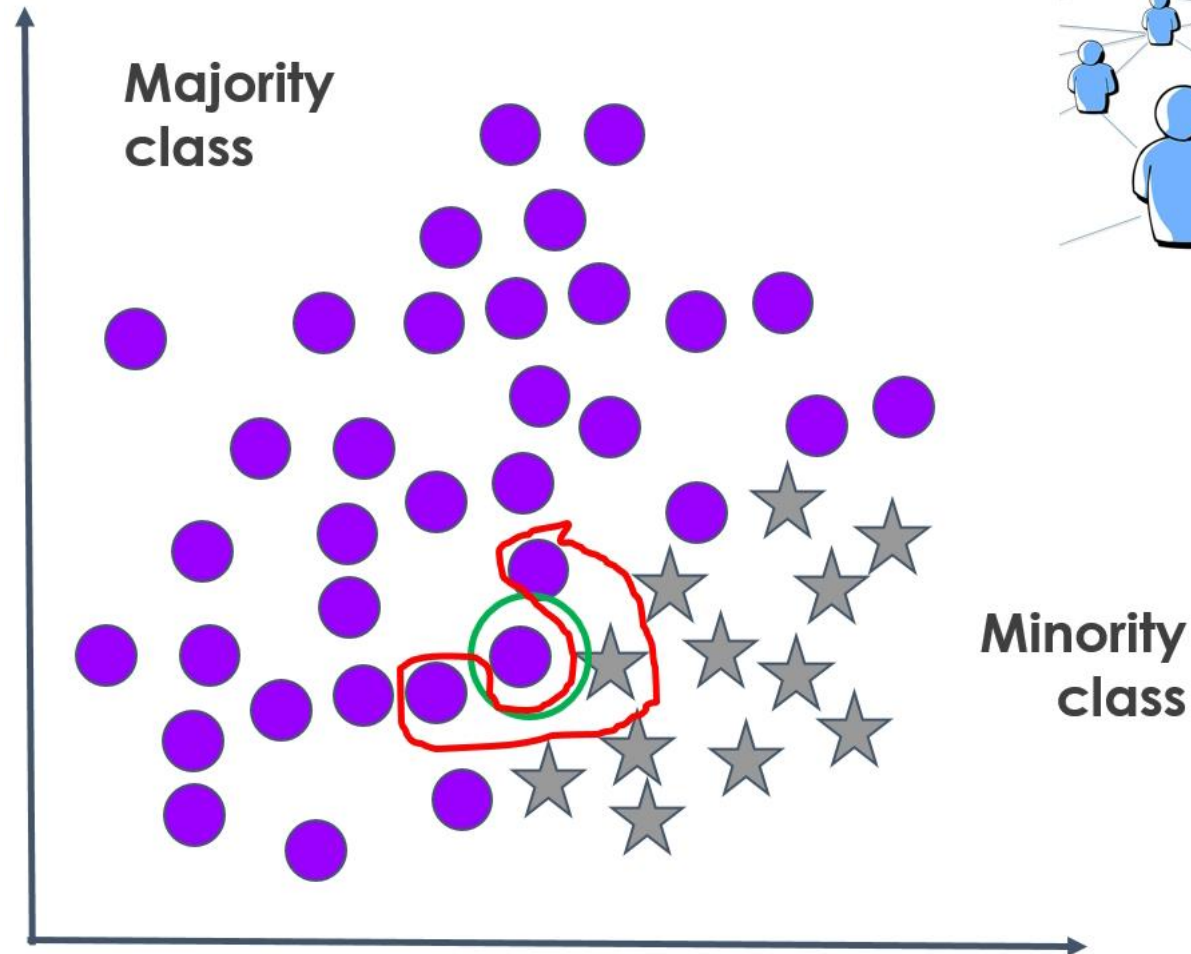
Edited Nearest Neighbours

If some neighbours show a different class, the observation is removed from the dataset.



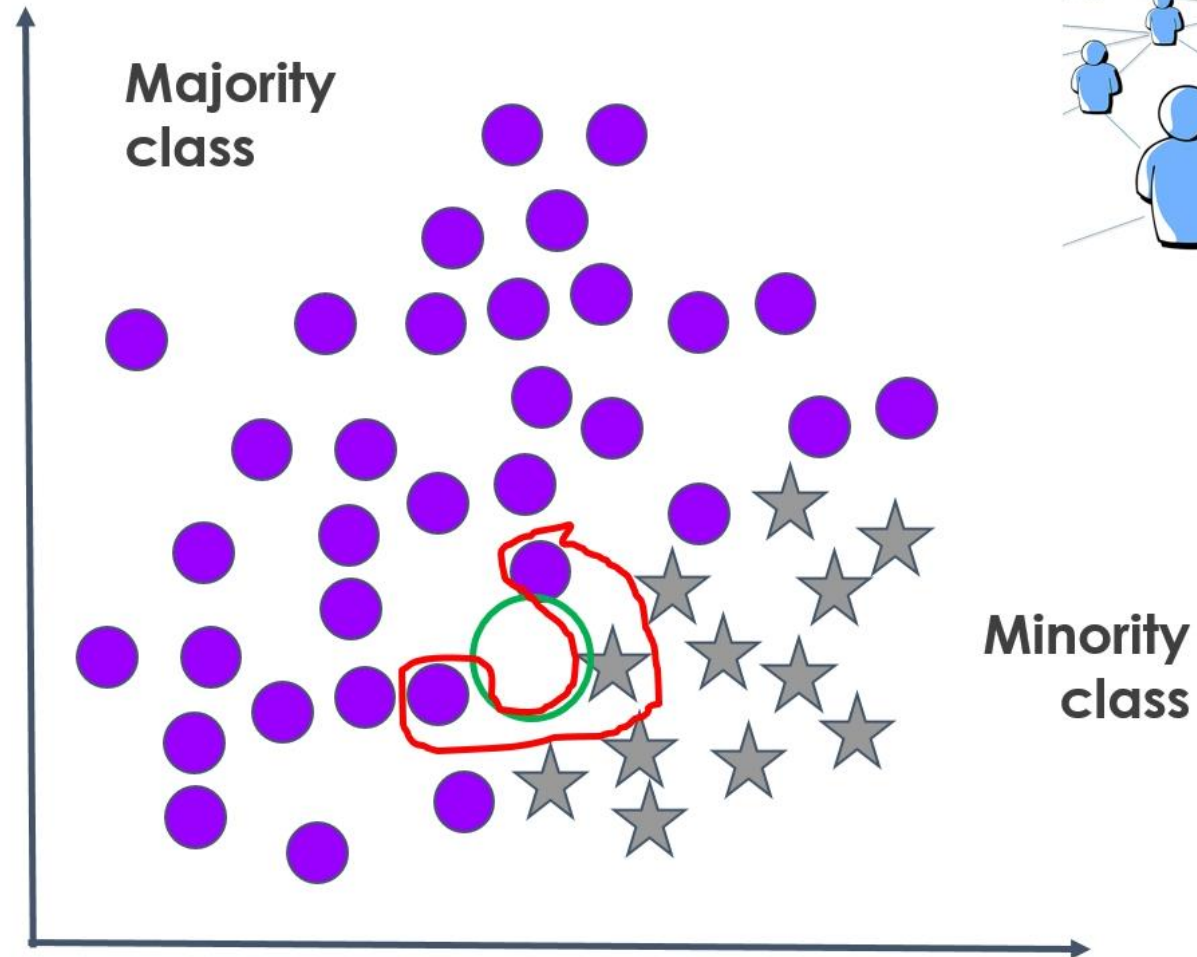
Edited Nearest Neighbours

If some neighbours show a different class, the observation is removed from the dataset.



Edited Nearest Neighbours

If some neighbours show a different class, the observation is removed from the dataset.



Imbalanced-learn: ENN

```
[3]: # create data

X, y = make_data(sep=2)

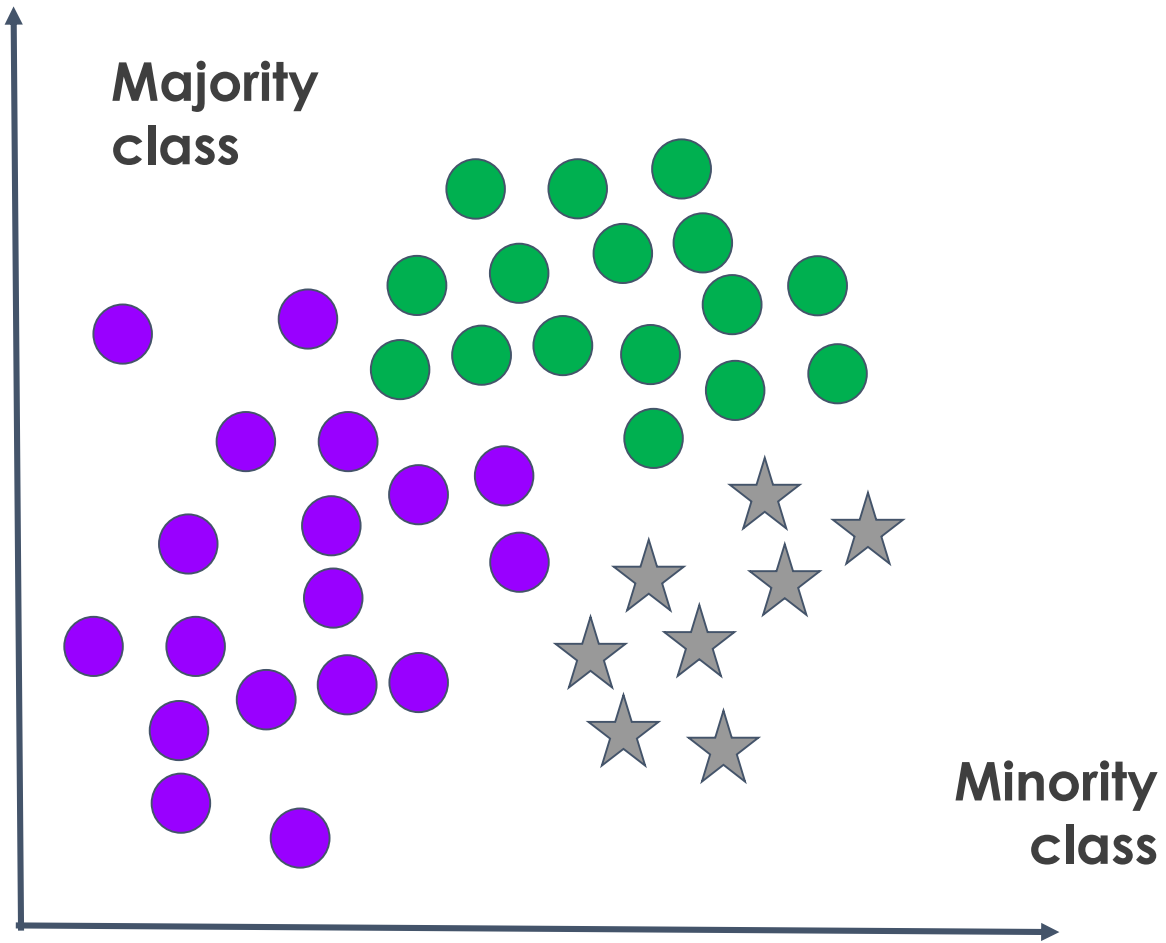
# set up edited nearest neighbour

enn = EditedNearestNeighbours(
    sampling_strategy='auto', # undersamples only the majority class
    n_neighbors=3,
    kind_sel='all', # all neighbours need to have the same label as the observation examined
    n_jobs=4) # I have 4 cores in my laptop

X_resampled, y_resampled = enn.fit_resample(X, y)
```

Number of neighbours to that will be examined.

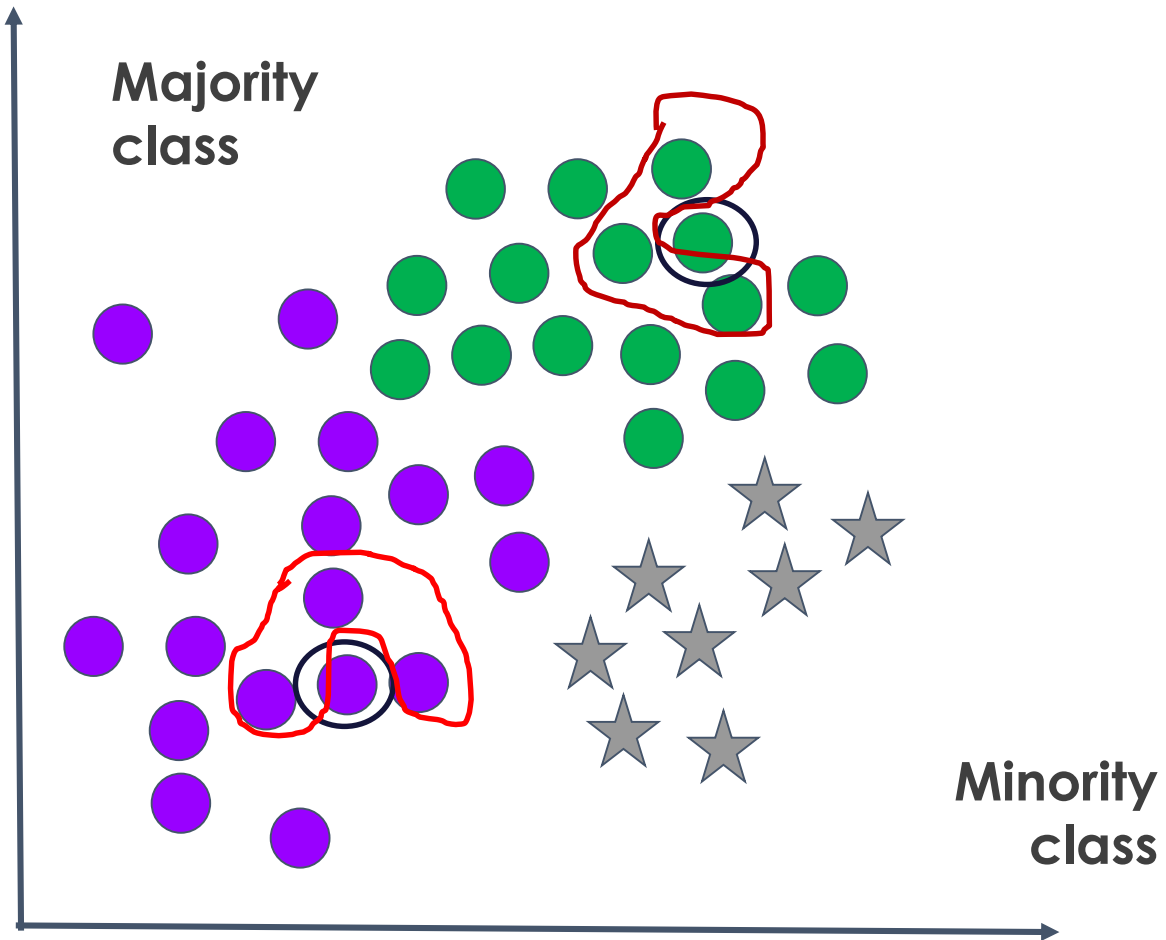
Multi-class



One vs Rest

Only majority classes are undersampled.

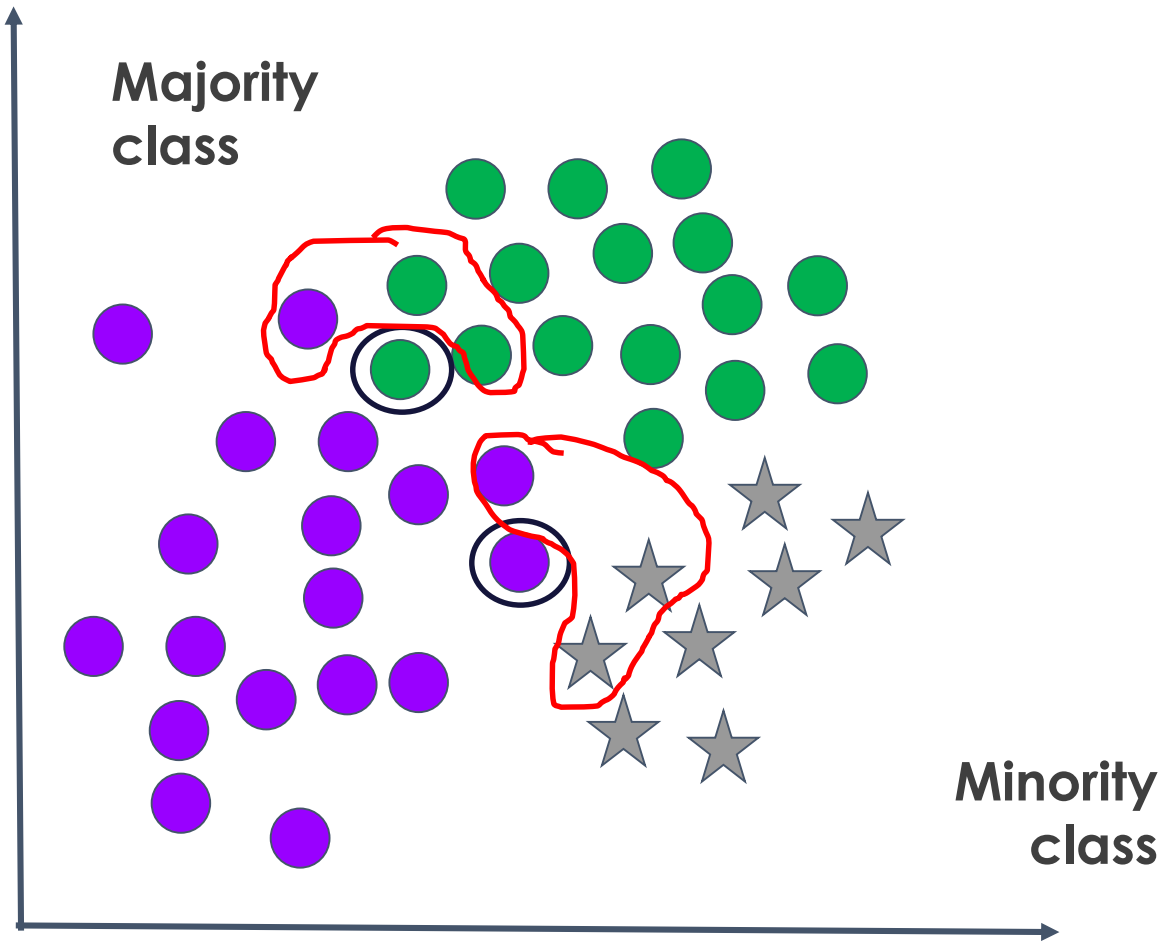
Multi-class



One vs Rest

When all or most neighbours agree, the observation is retained.

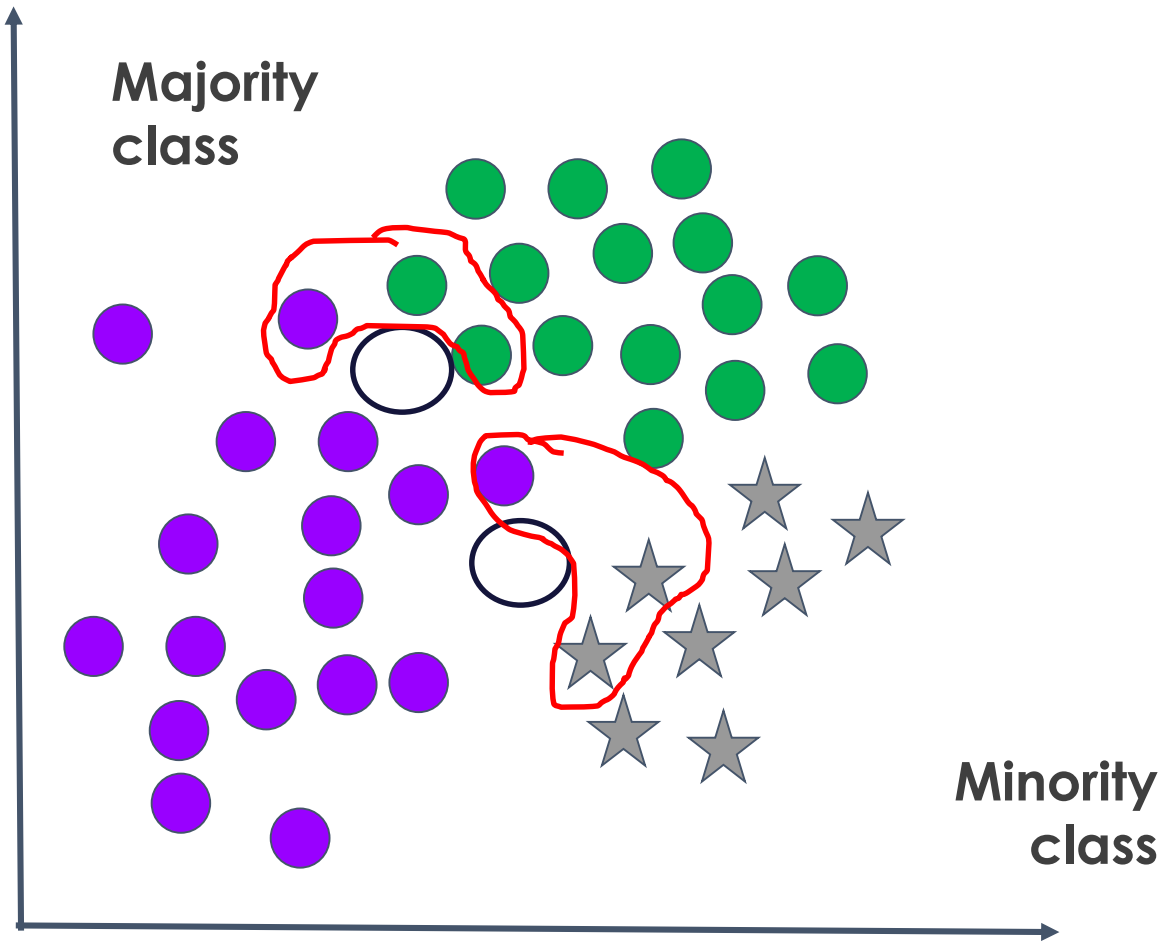
Multi-class



One vs Rest

When all or most neighbours agree, the observation is retained.

Multi-class



One vs Rest

When neighbours disagree,
the observation is
removed.

THANK YOU

www.trainindata.com