Train In Data

Repeated Edited Nearest Neighbours

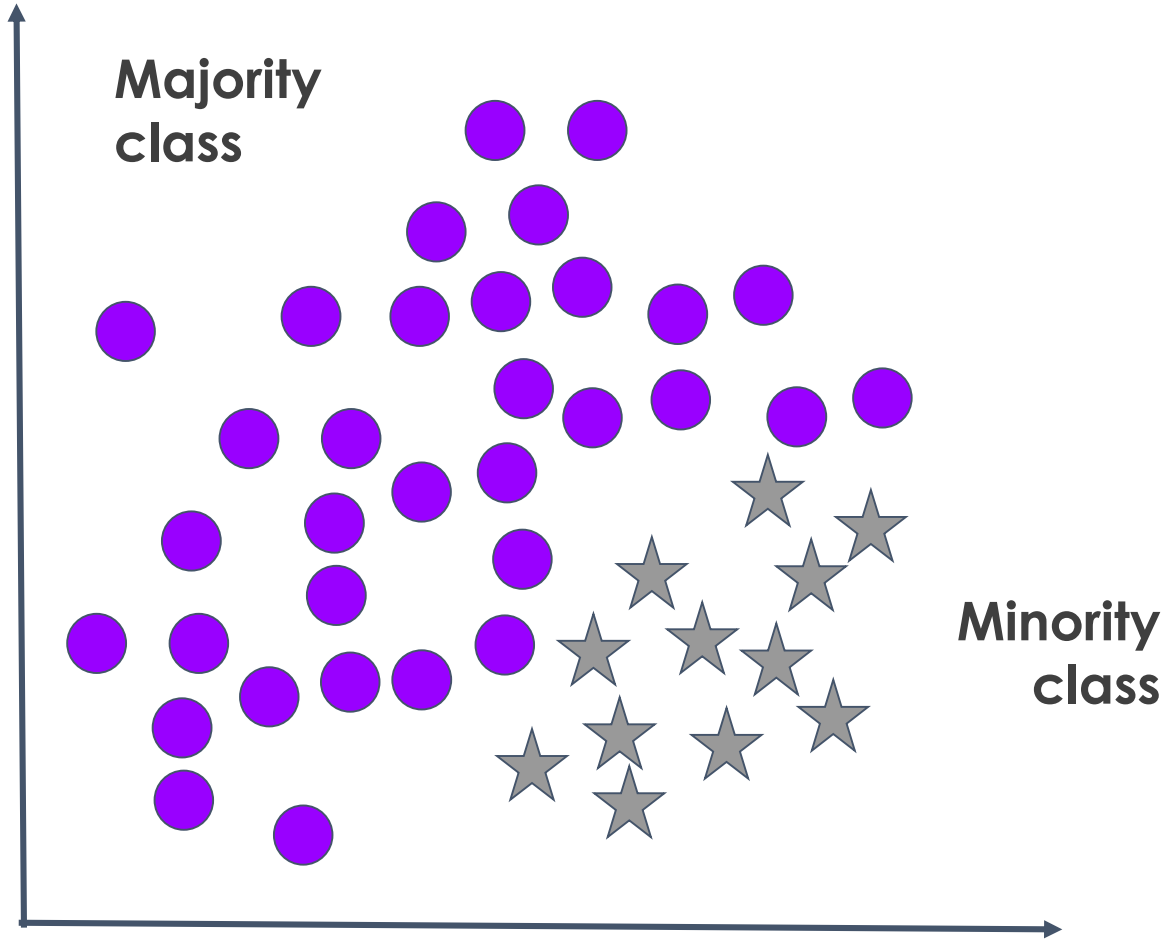
Repeated Edited Nearest Neighbours

1. Trains 3 KNN on entire dataset
2. Finds each observation's 3 neighbours
3. Decides whether to keep or remove, based on neighbours agreement with its class.
4. Repeats
    - Until no more observations are removed
    - A maximum number of cycles is reached

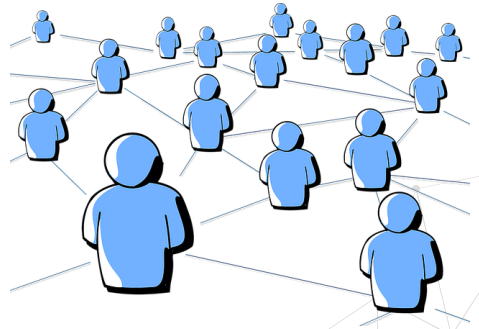
Repeated Edited Nearest Neighbours

- Final dataset shape varies
- Cleaning
- Removes hard cases

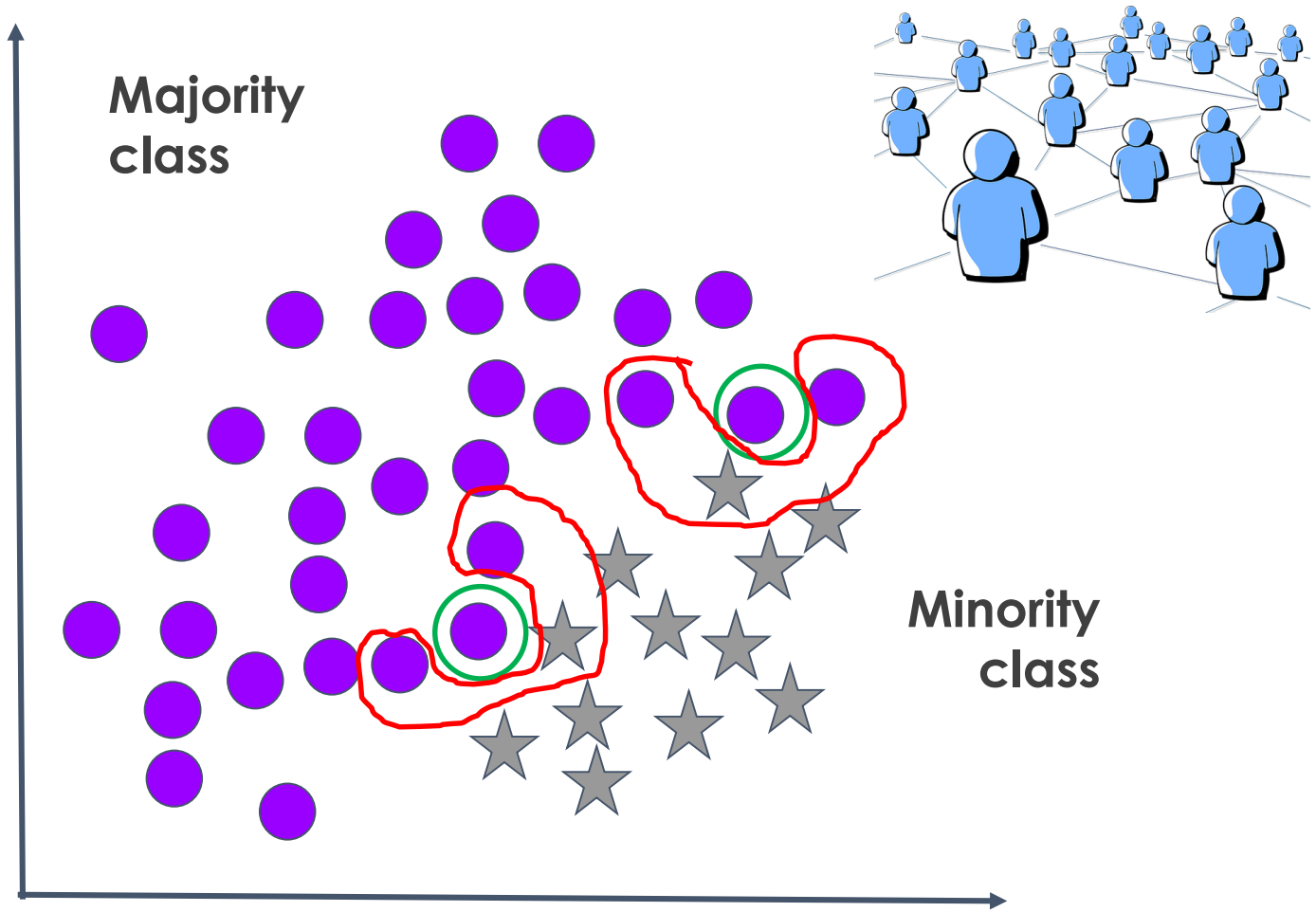
Edited Nearest Neighbours



Train a 3 KNN algorithm

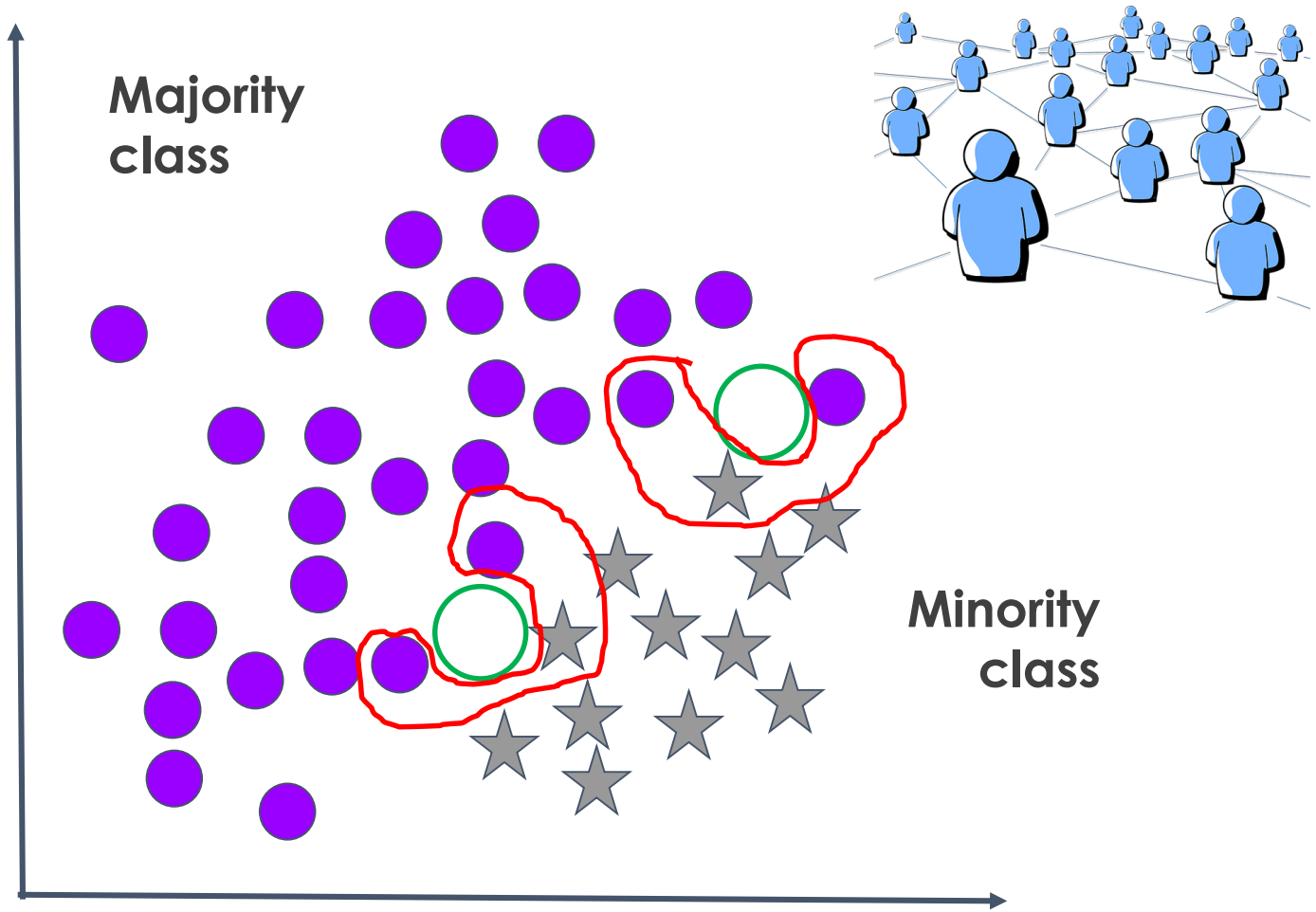


Edited Nearest Neighbours



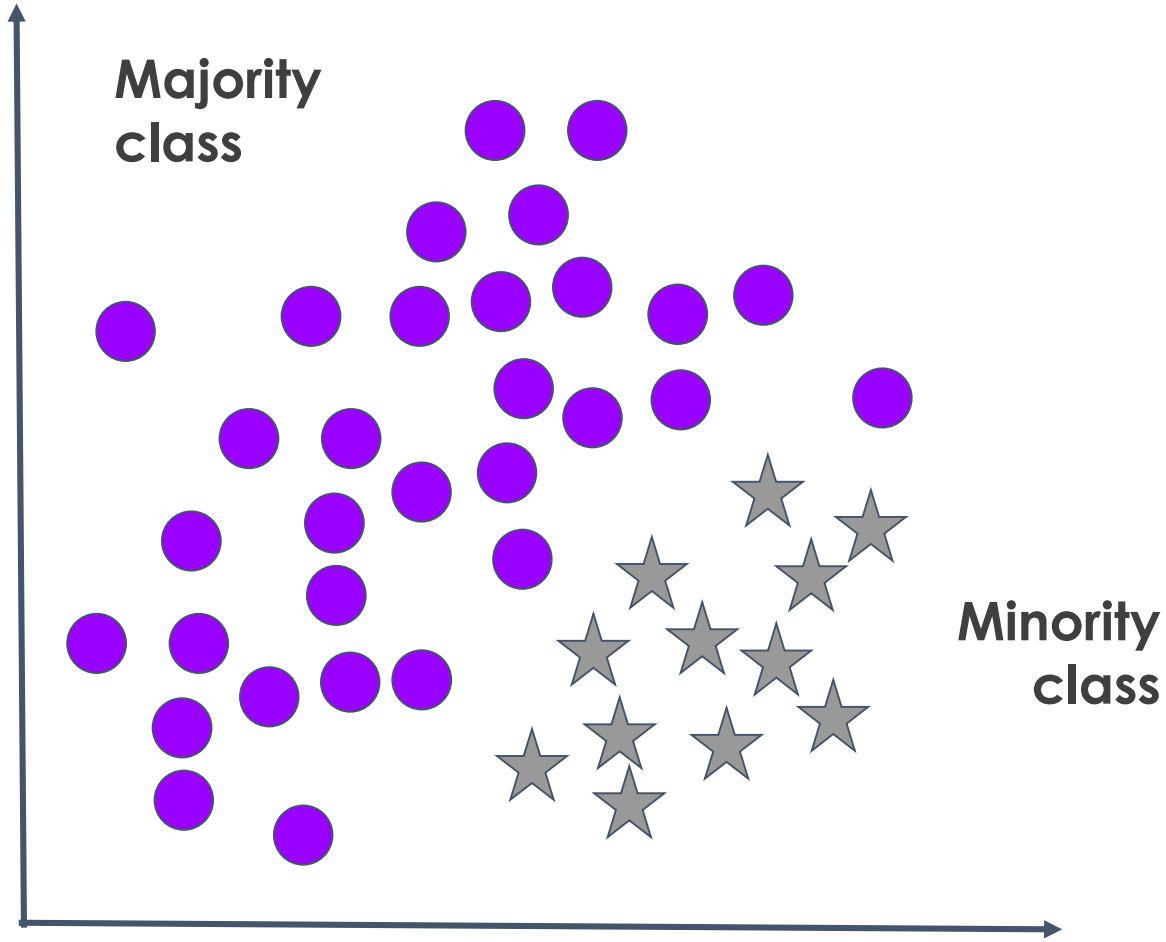
- Find each observations 3 closest neighbours.
- For simplicity, in the diagram I only show those where some of the neighbours disagree with the class.

Edited Nearest Neighbours

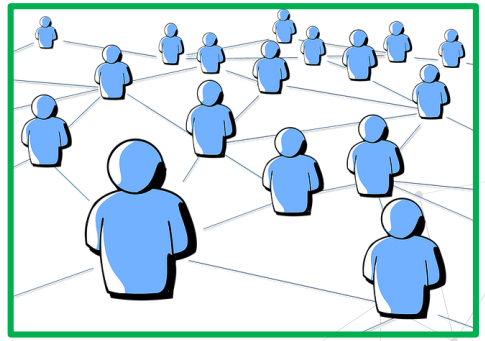


- Remove observations from the majority class, when neighbours disagree.

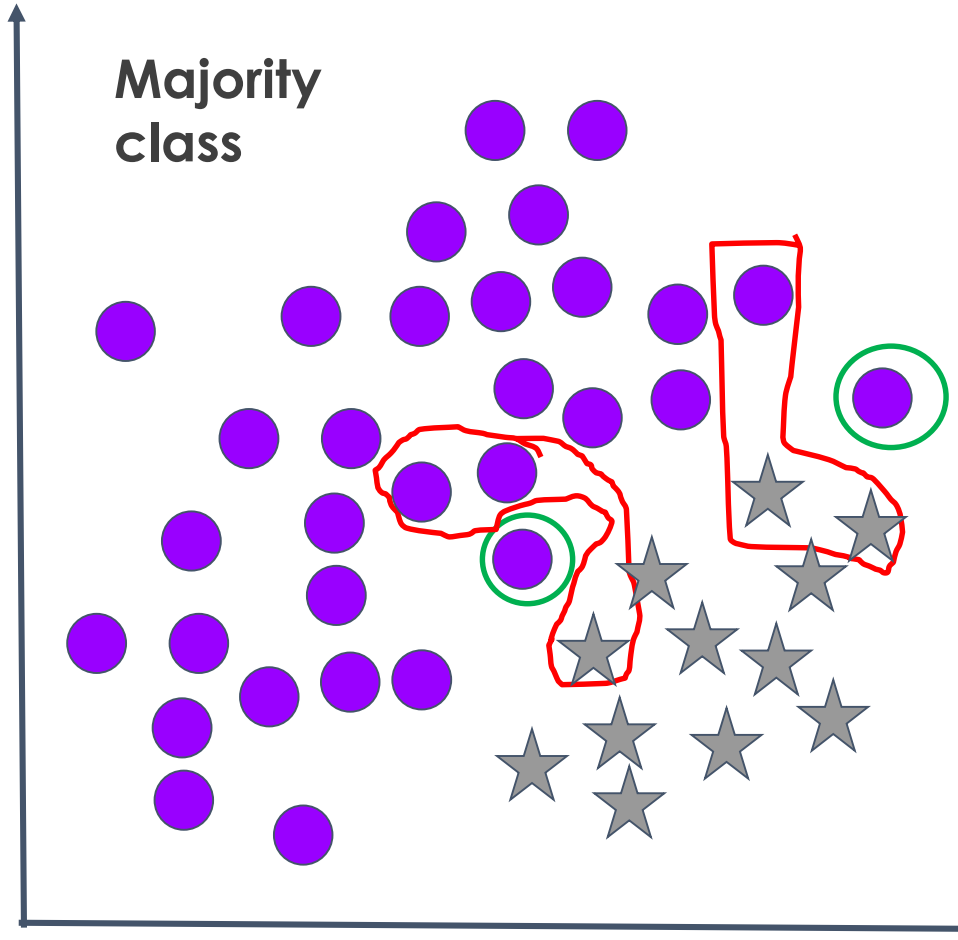
Repeated Edited Nearest Neighbours

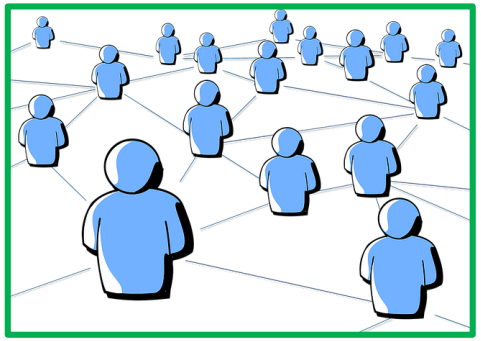


Train a new 3 KNN algorithm



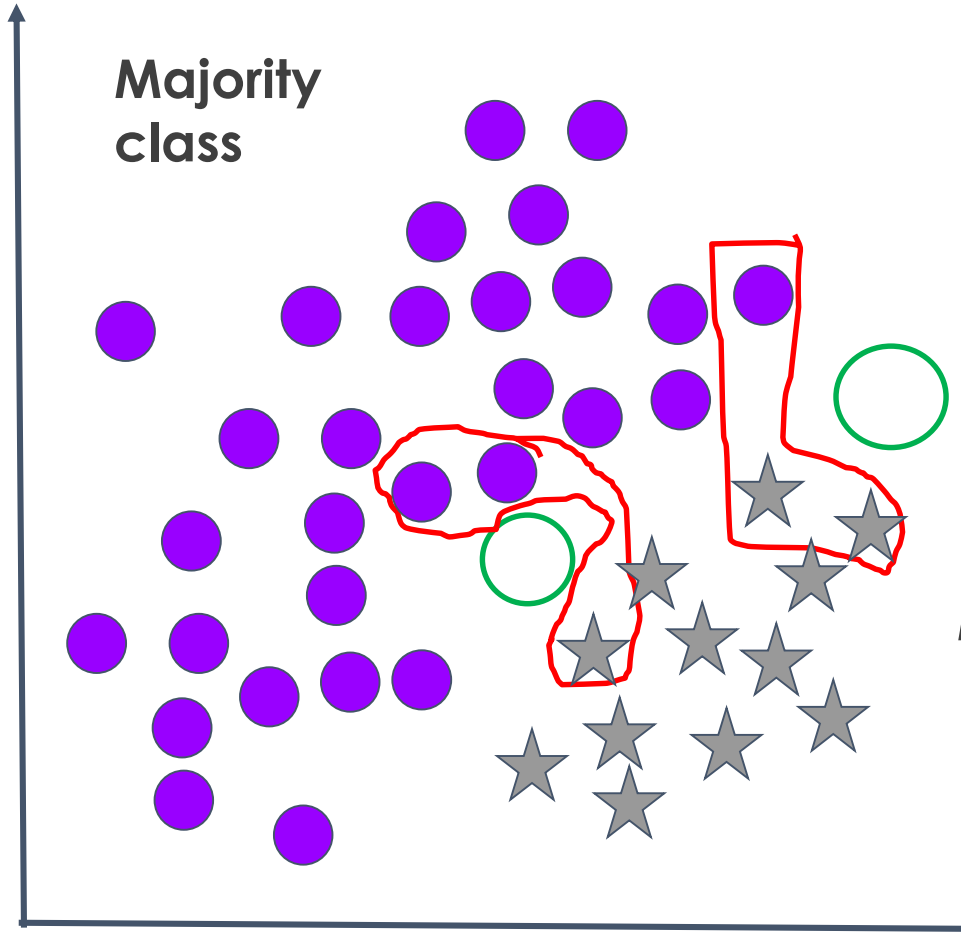
Repeated Edited Nearest Neighbours

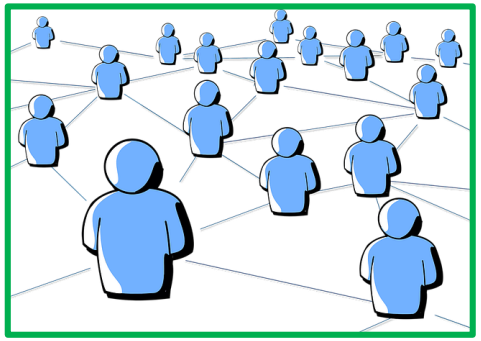




- Find each observations 3 closest neighbours.
- For simplicity, in the diagram I only show those where some of the neighbours disagree with the class.

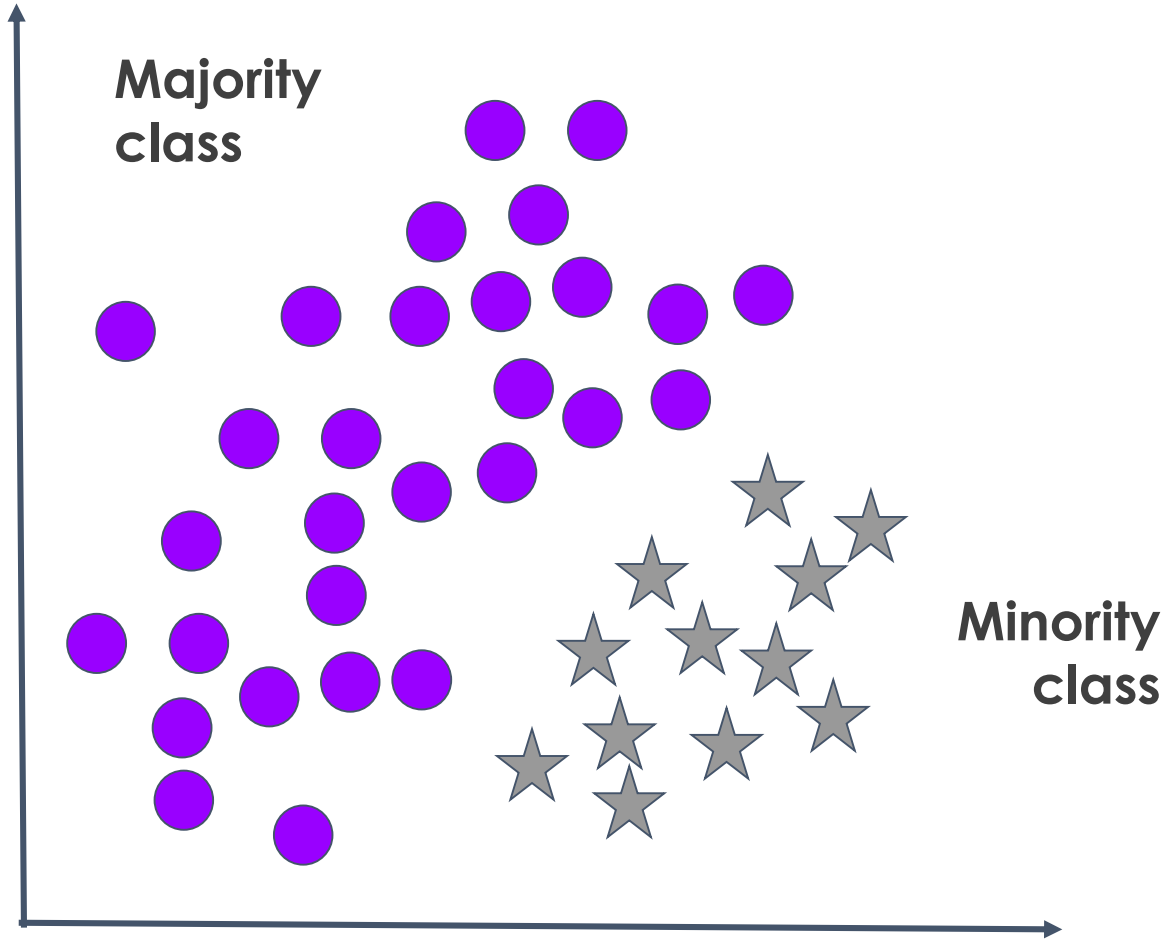
Repeated Edited Nearest Neighbours



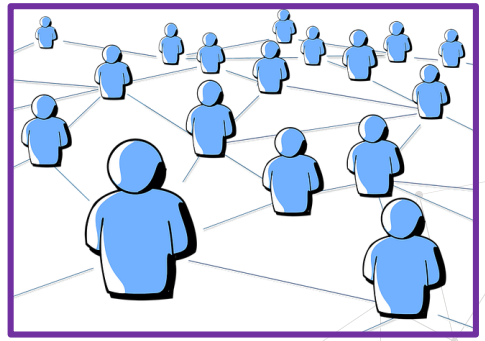


- Remove observations from the majority class, when neighbours disagree.

Repeated Edited Nearest Neighbours



Train another 3 KNN algorithm



Repeated Edited Nearest Neighbours

Repeat N times.

Repeated Edited Nearest Neighbours

1. Removes more samples than ENN
2. Various passes over the dataset
3. Always builds KNN with same number of neighbours
    - 3 KNN

Imbalanced-learn: RENN

:

```
# create data

X, y = make_data(sep=2)

# set up repeated edited nearest neighbour

renn = RepeatedEditedNearestNeighbours(
    sampling_strategy='auto',# removes only the majority class
    n_neighbors=3, # 3 KNN
    kind_sel='all', # all neighbouring observations should show the same class
    n_jobs=4, # 4 processors in my laptop
    max_iter=100) # maximum number of iterations

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

Multi-class

One vs Rest

Train In Data

THANK YOU

www.trainindata.com