Train In Data

Random Under-sampling

Random under-sampling

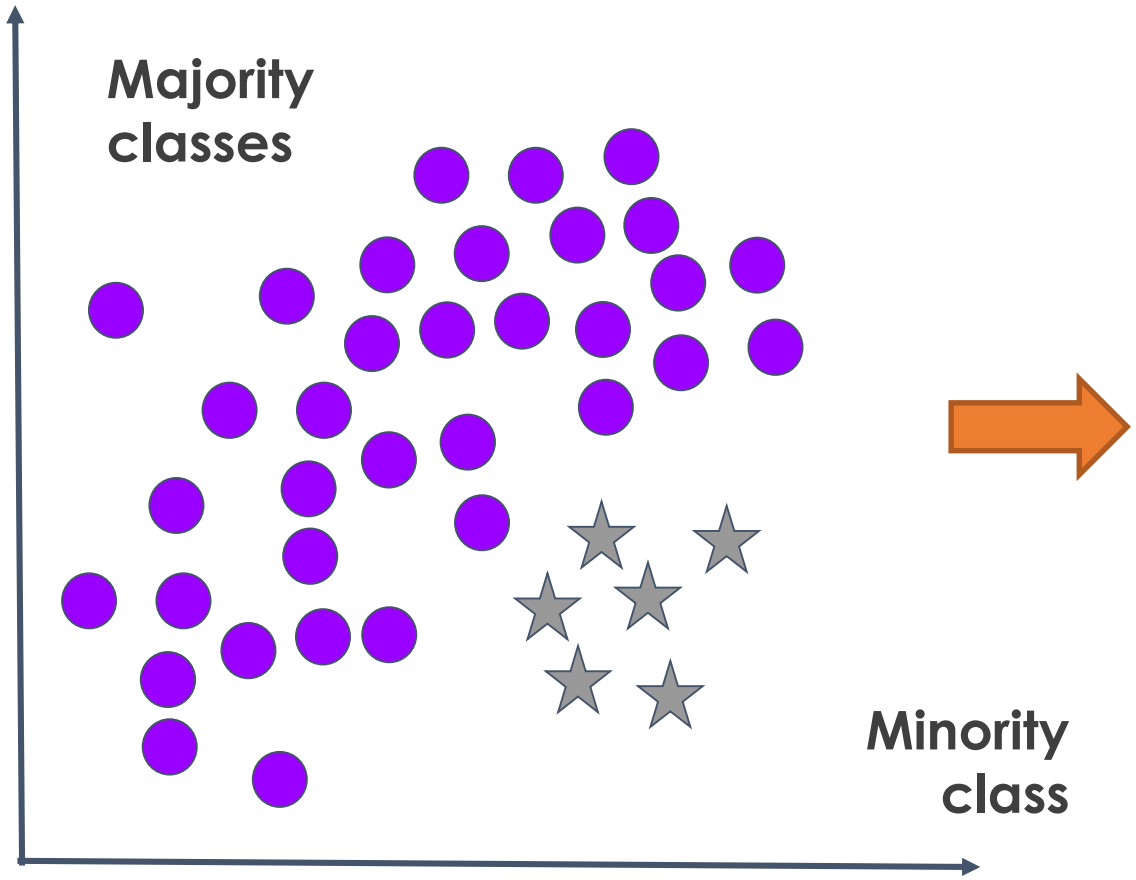
Extracts observations at random from the majority class, until a certain balancing ratio is reached.

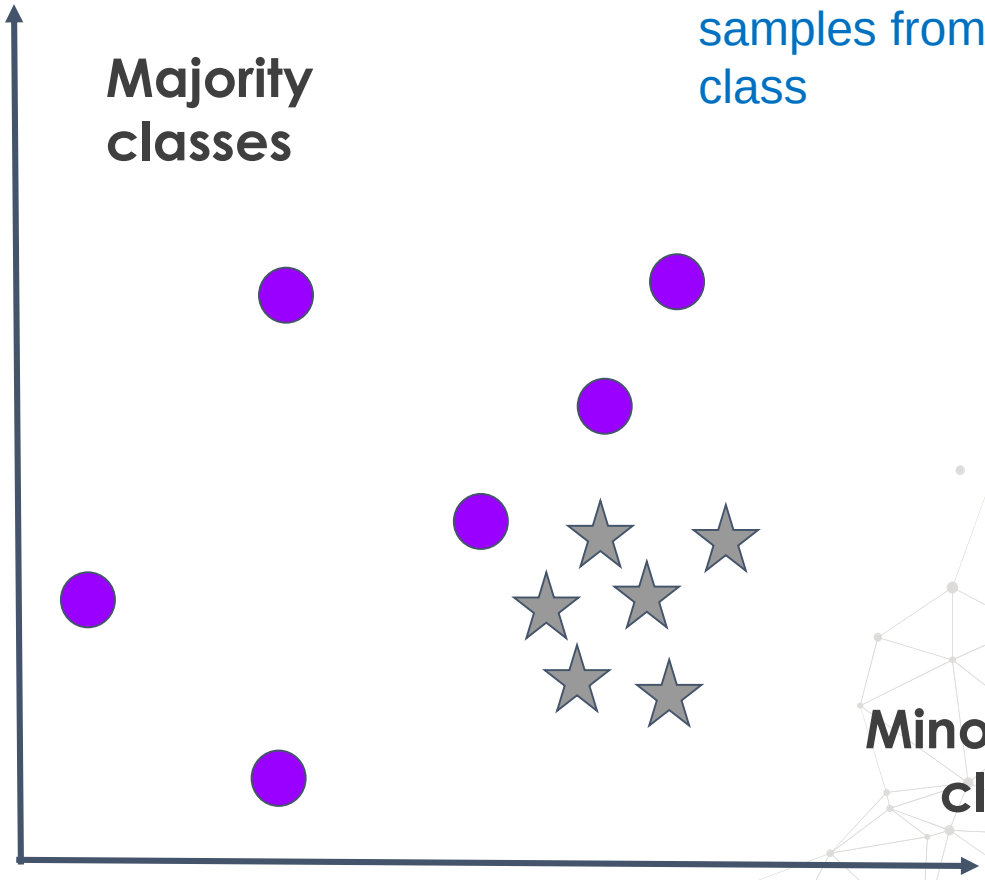
Naïve technique.

Random under-sampling

Balancing ratio = 1

Equal number of samples from each class

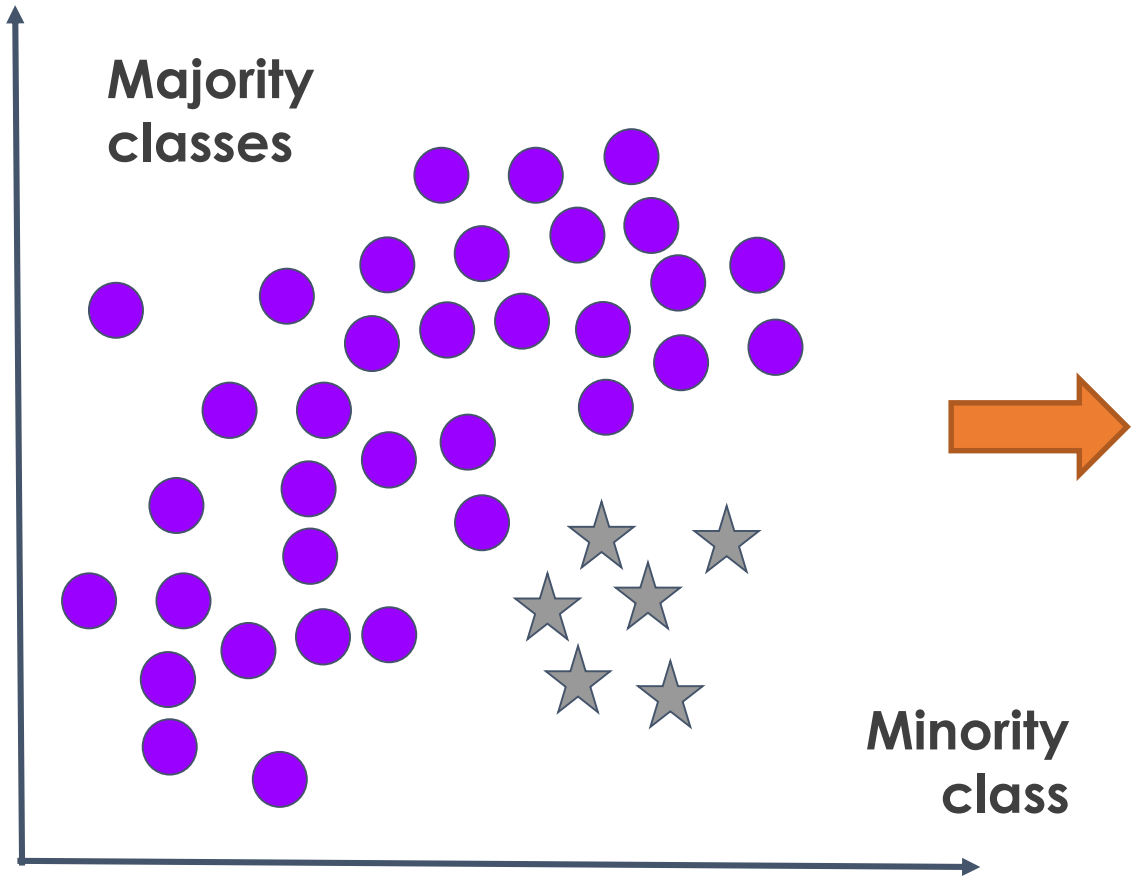


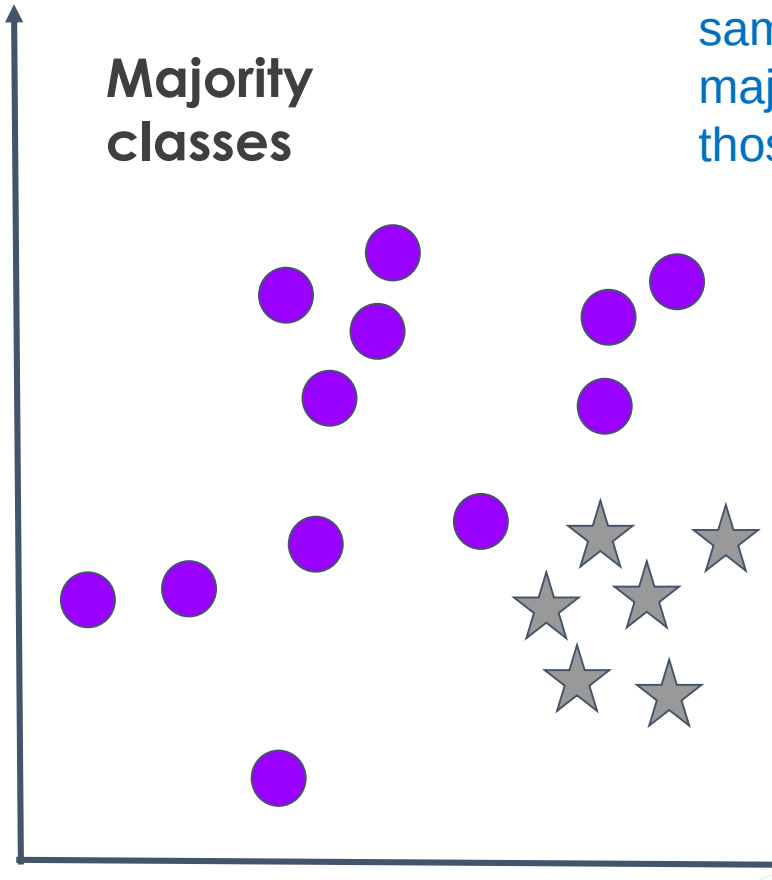


Random under-sampling

Balancing ratio = **0.5**

Twice as many samples from majority classes as those from minority.

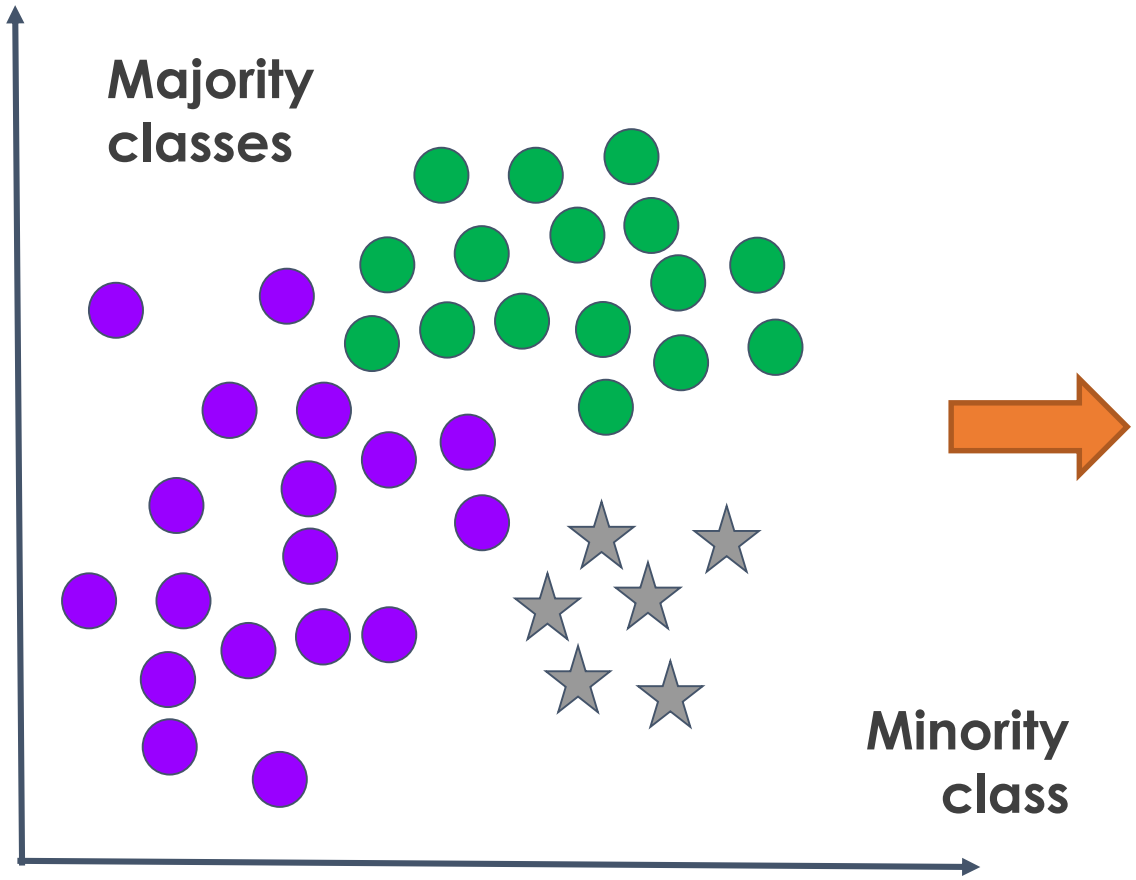


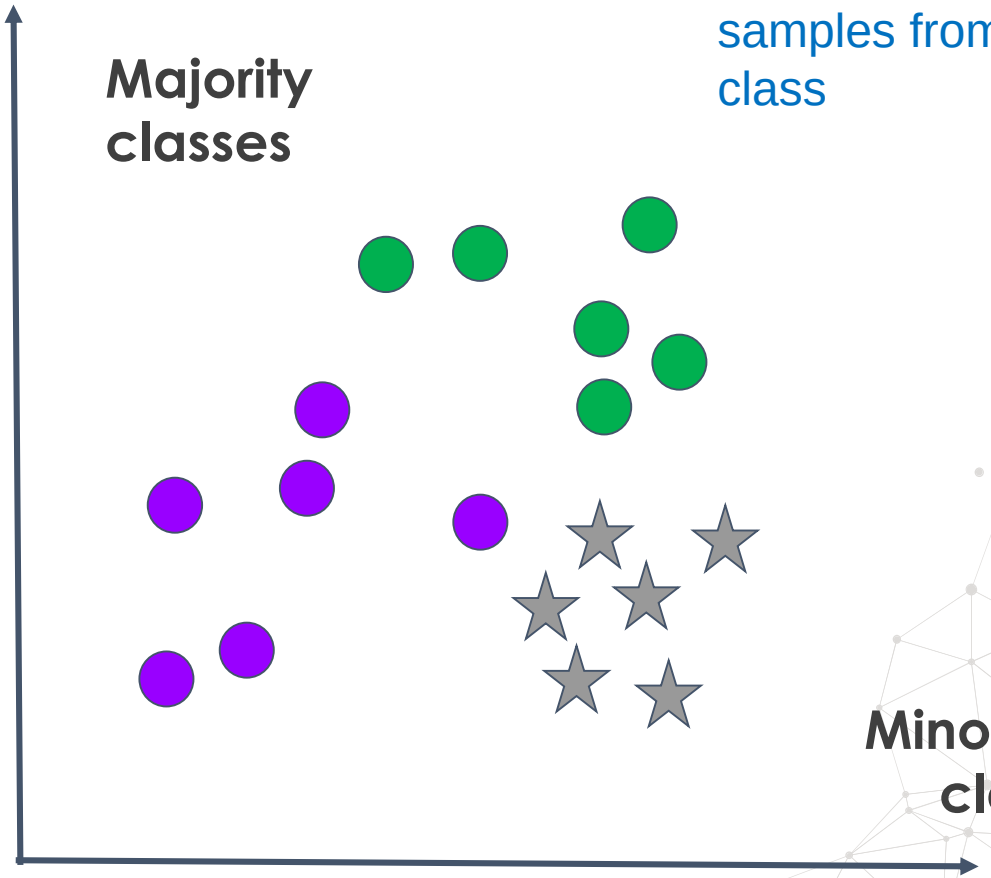


Multi-class

Balancing ratio = 1

Equal number of samples from each class





Imbalanced-learn: RUS

sampling_strategy:

Imbalanced-learn will evaluate the target and determine the class with least observations as the minority. Then it will undersample all other classes.

```
# create data

X, y = make_data(sep=2)

# set up the random undersampling class

rus = RandomUnderSampler(
    sampling_strategy='auto',  # samples only the majority class
    random_state=0,  # for reproducibility
    replacement=True # if it should resample with replacement
)

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

Imbalanced-learn: RUS

sampling_strategy:

If multiclass with more than 1 minority class, or we want to attain a different balancing ratio, we need to specify it.

:

```
# create data

X, y = make_data(sep=2)

# set up the random undersampling class

rus = RandomUnderSampler(
    sampling_strategy= {1: 300, 2:300, 3:300}
    random_state=0,  # for reproducibility
    replacement=True # if it should resample with replacement
)

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

Imbalanced-learn: RUS

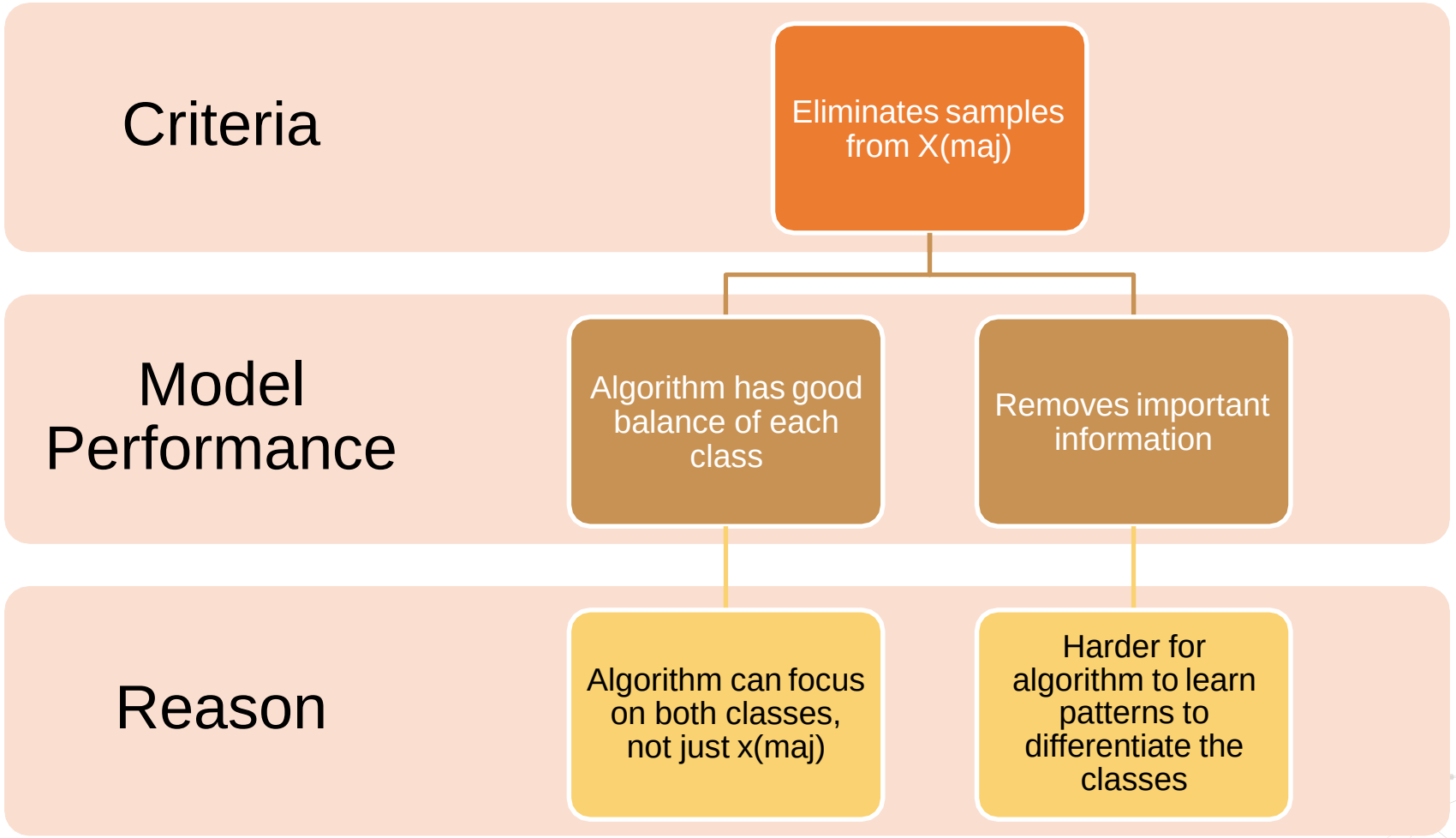
:

```
# create data

X, y = make_data(sep=2)

# set up the random undersampling class

rus = RandomUnderSampler(
    sampling_strategy='auto',  # samples only the majority class
    random_state=0,  # for reproducibility
    replacement=True # if it should resample with replacement
)

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

replacement:

If True, the same observation can be sampled more than once, in general False, unless we have very few observations.

RUS: considerations



Train In Data

THANK YOU

www.trainindata.com