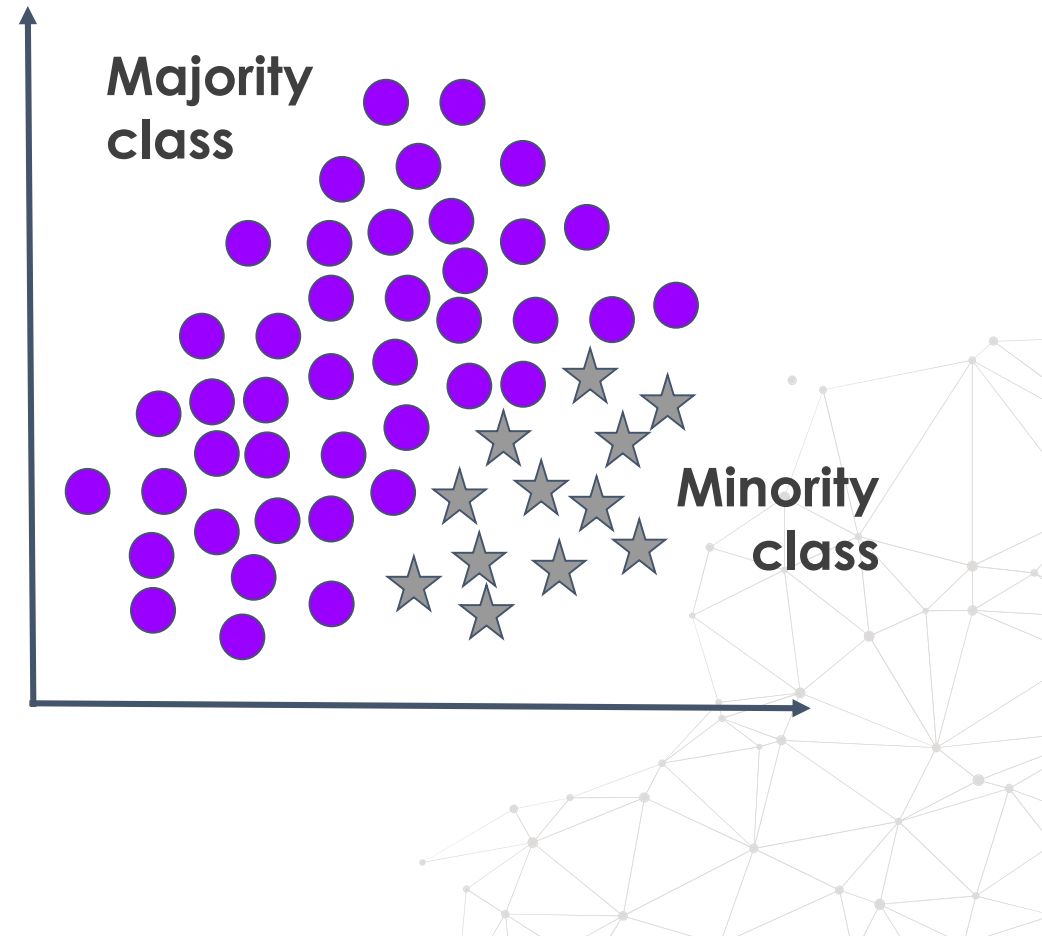




Under-sampling Methods

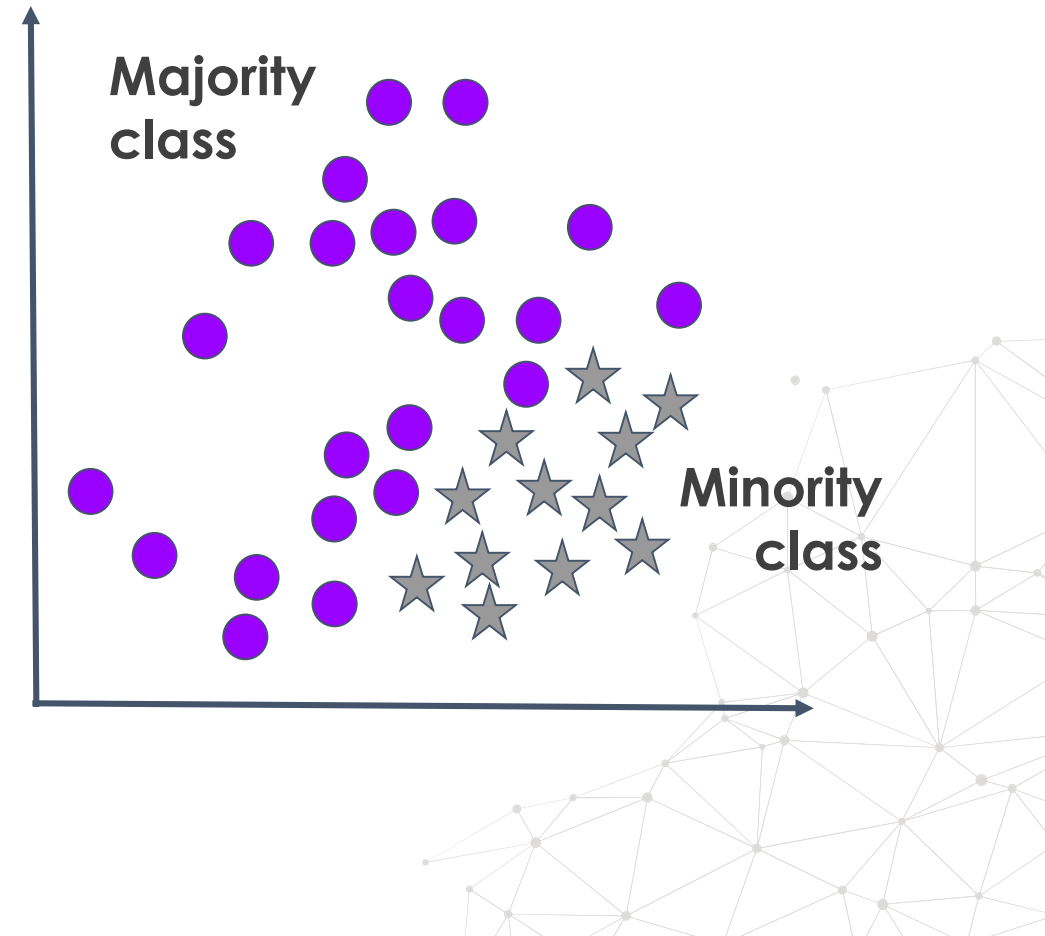
Under-sampling methods

Process of **reducing** the number of samples from the Majority Class

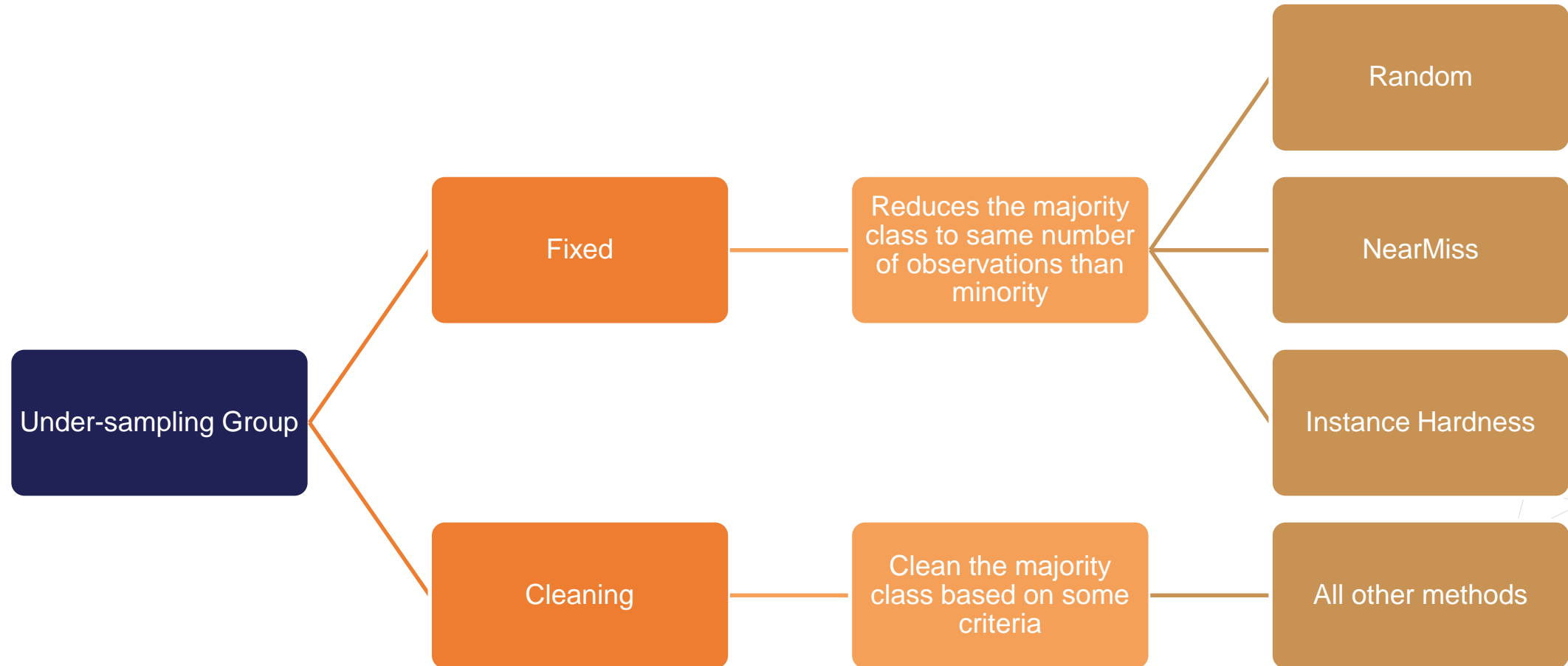


Under-sampling methods

Process of **reducing** the number of samples from the Majority Class



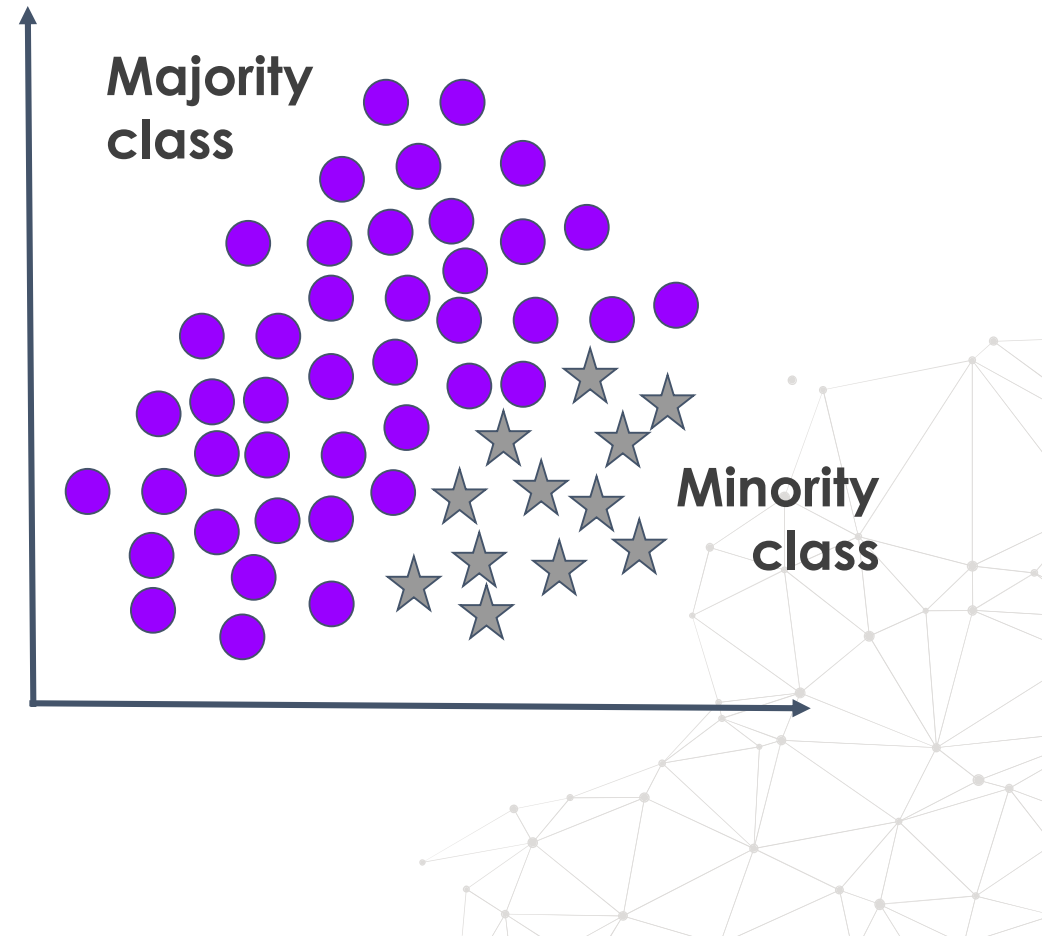
Fixed vs Cleaning under- sampling



Balancing ratio

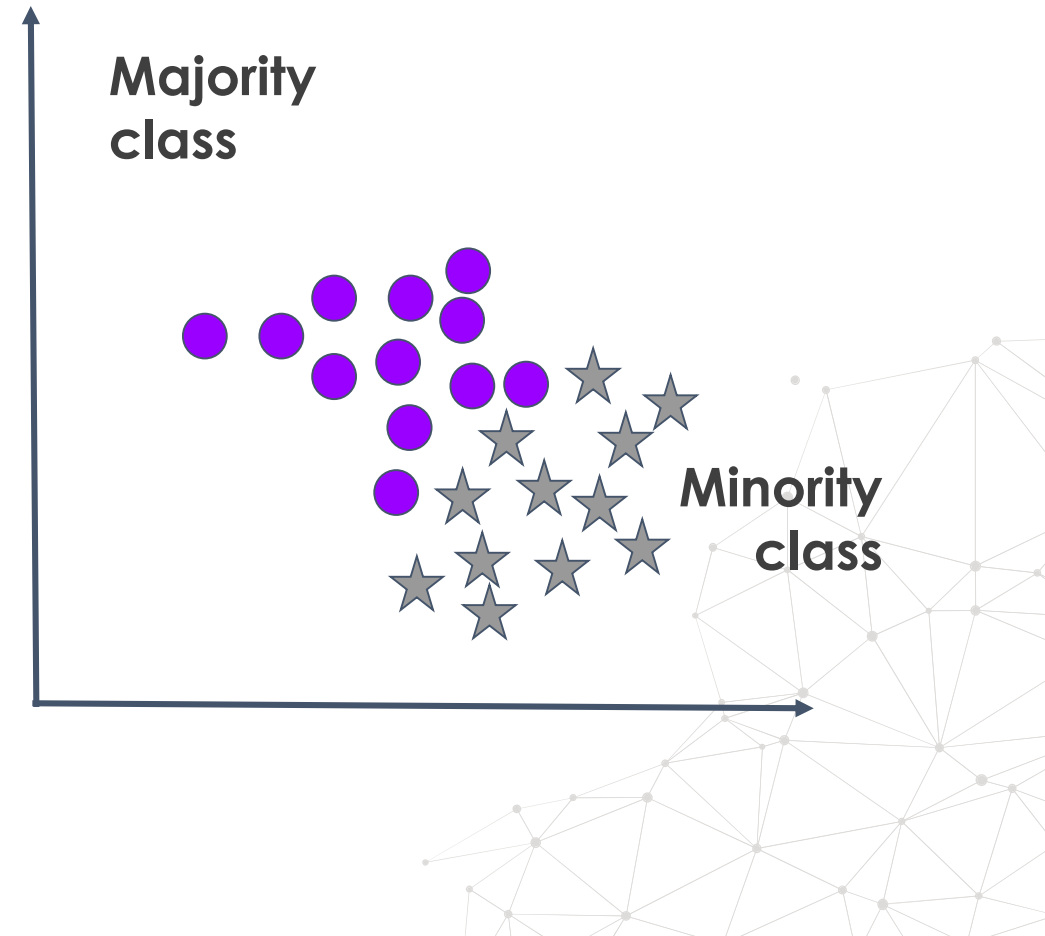
$$R(X) = \frac{X(\text{minority})}{X(\text{majority})}$$

- If $X(\text{min}) = X(\text{maj}) \rightarrow R(x) = 1$
- If $X(\text{min}) = 1/2 \times X(\text{maj}) \rightarrow R(x) = 0.5$



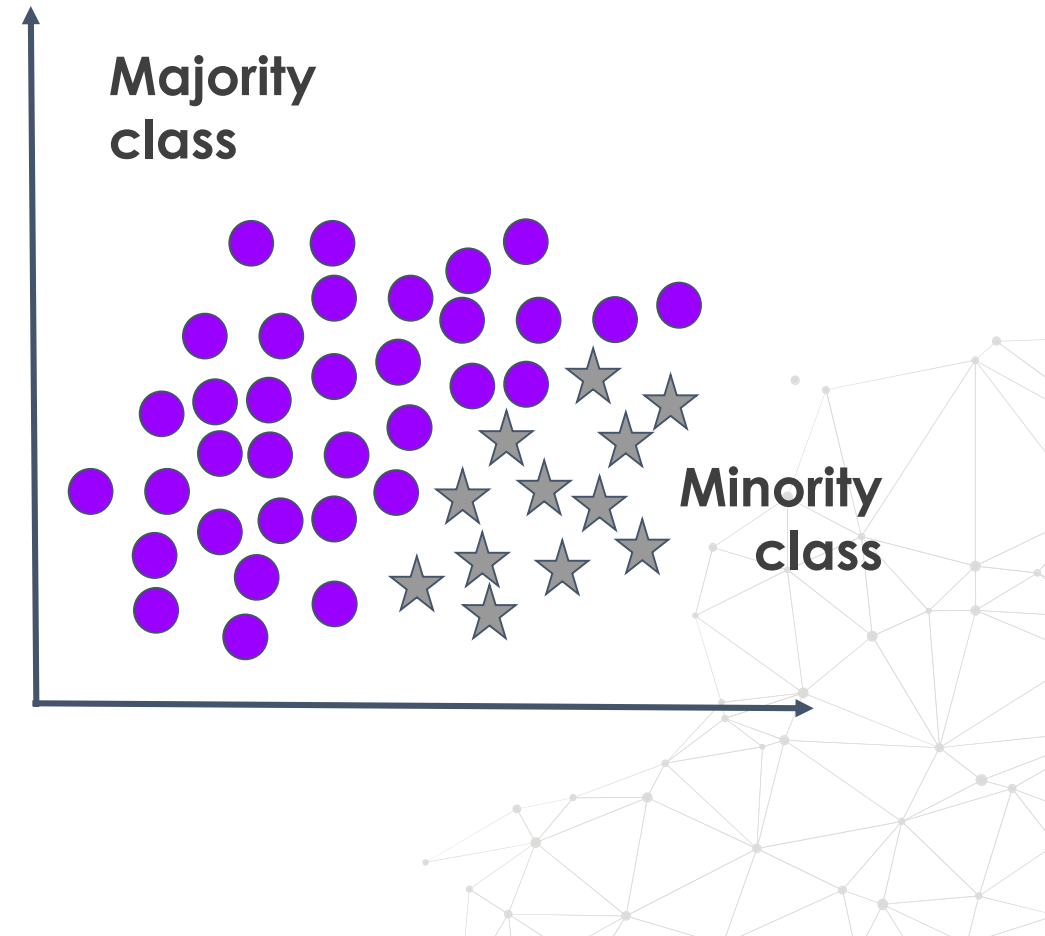
Fixed Under-sampling

- Remove samples from the majority class until $R(x) = 1$

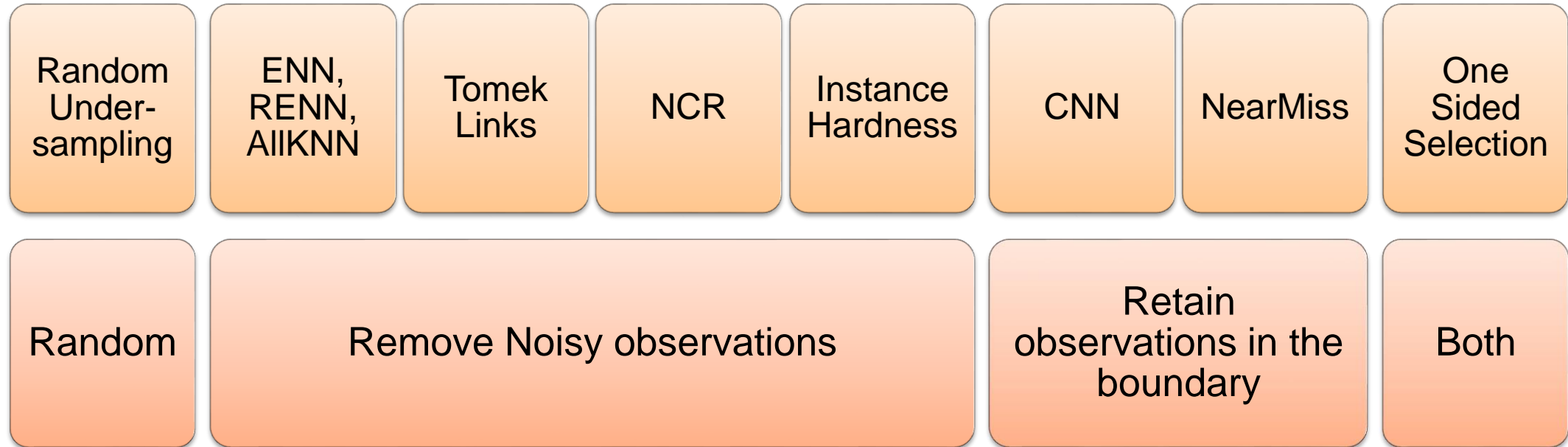


Fixed Under-sampling

- Remove samples from the majority class until $R(x) = 1$
- But, user could determine otherwise, for example $R(x) = 2$, that is twice as many from the majority class as those from the minority.

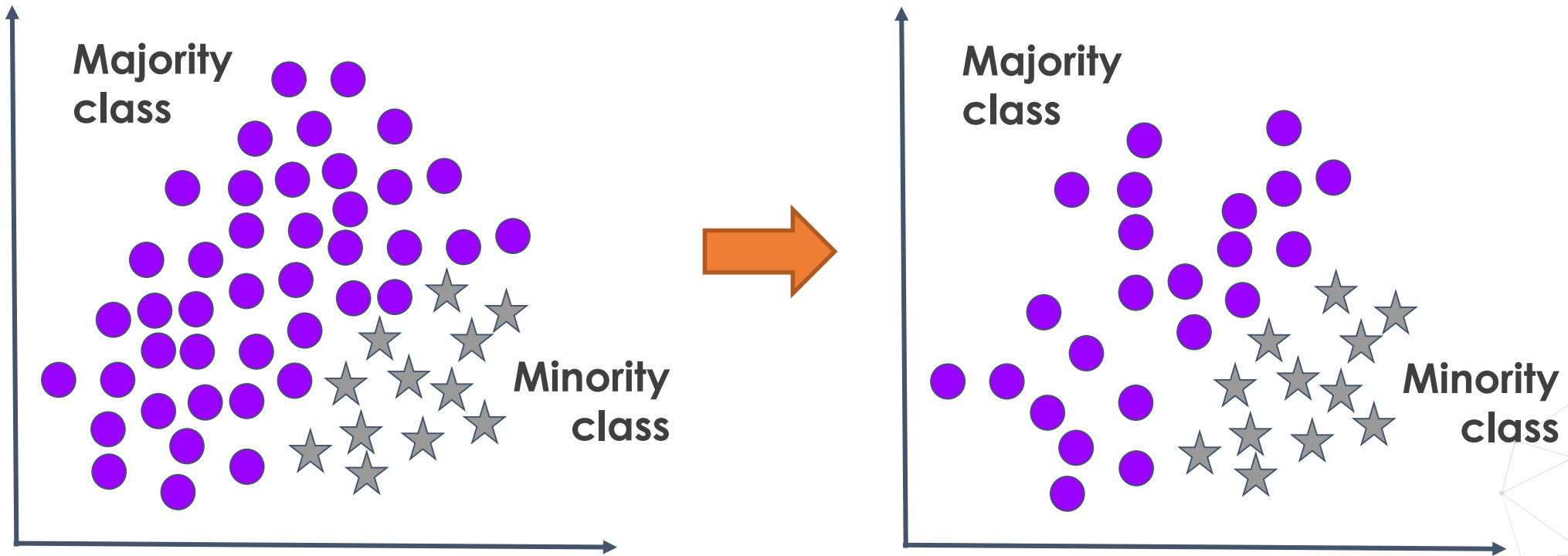


Under-sampling criteria



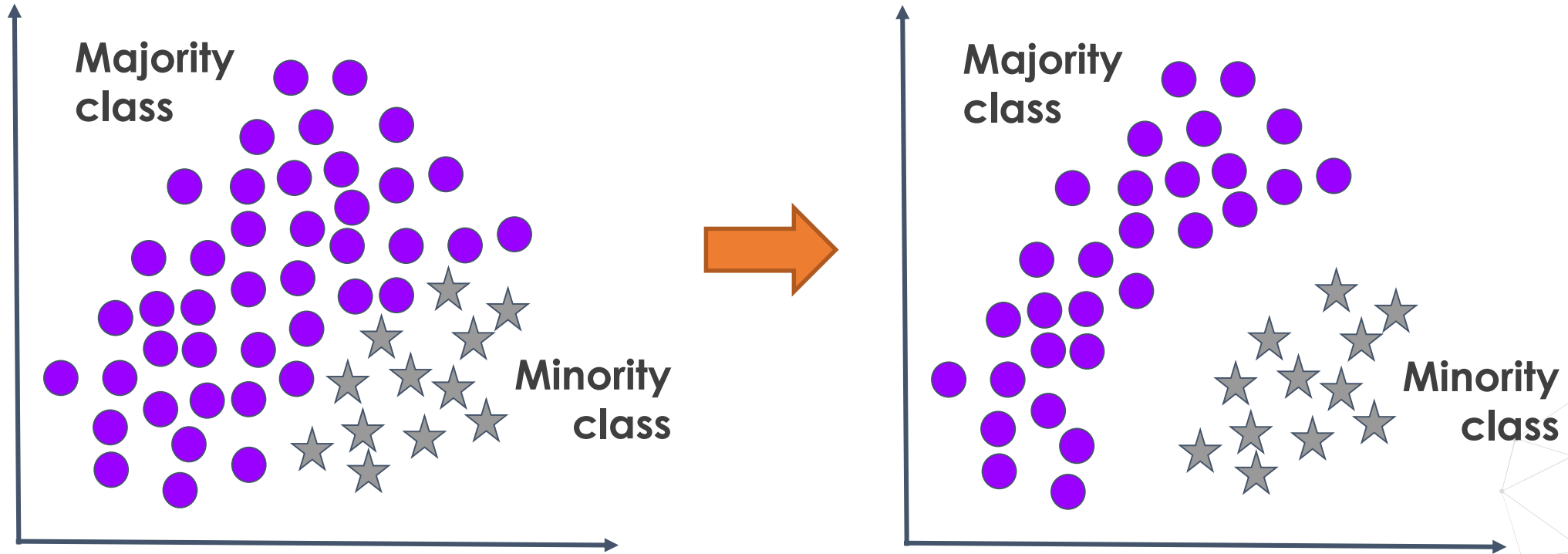
Under-sampling criteria

Random under-sampling



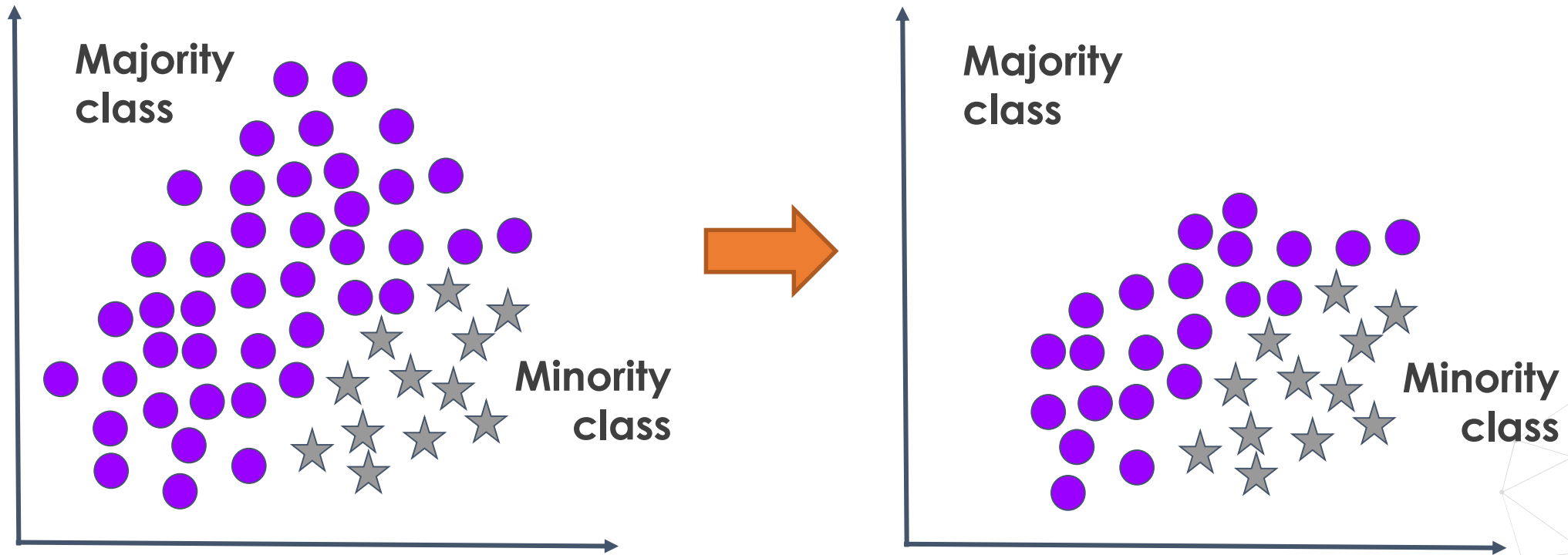
Observations are removed at random.

Remove noisy observations



Observations similar to the opposite class are preferentially removed.

Retain closer observations



Observations similar to the opposite class are preferentially retained.

THANK YOU

www.trainindata.com