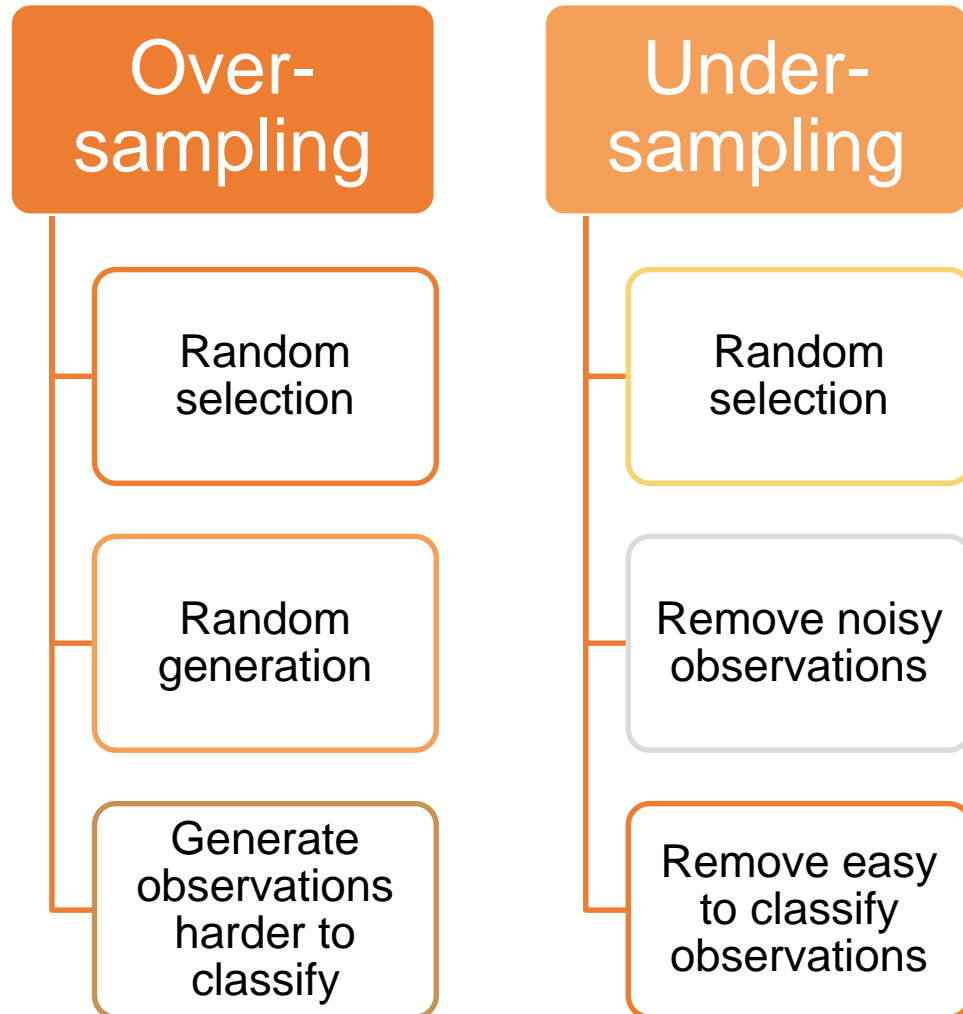


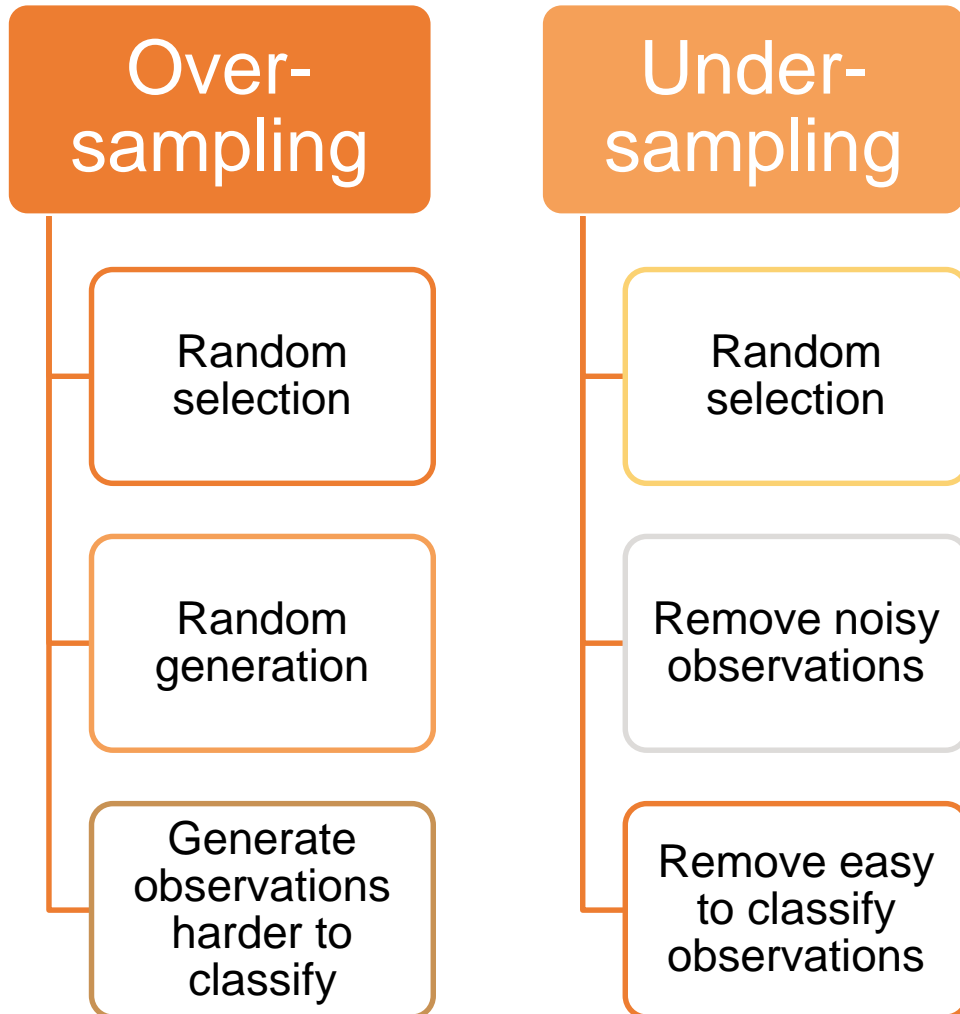


# Combining Over- and Under-sampling Methods

# Over- and under-sampling methods



# Over- and under-sampling: pros & cons



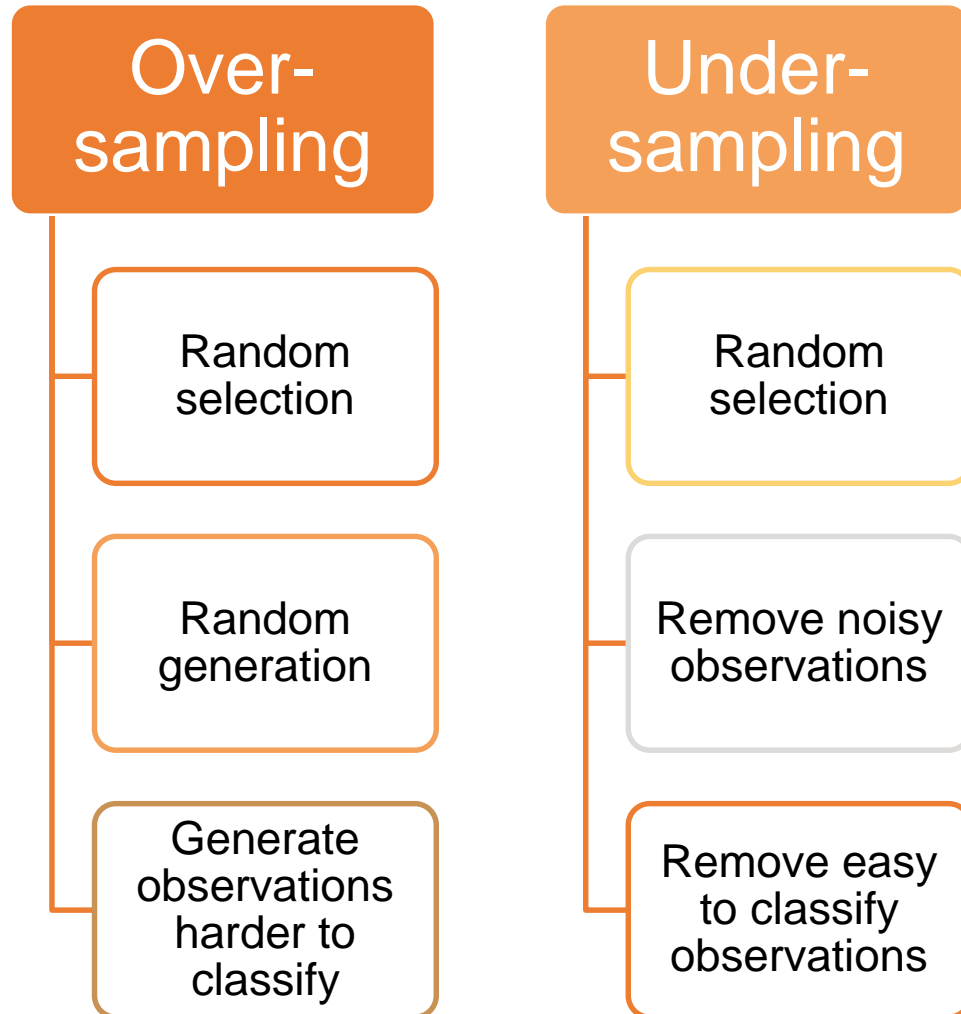
## Over-sampling

- More observations from minority class.
- Can amplify noise

## Under-sampling

- Lose important information (from majority class)
- Remove noisy observations

# Over- and under-sampling: pros & cons



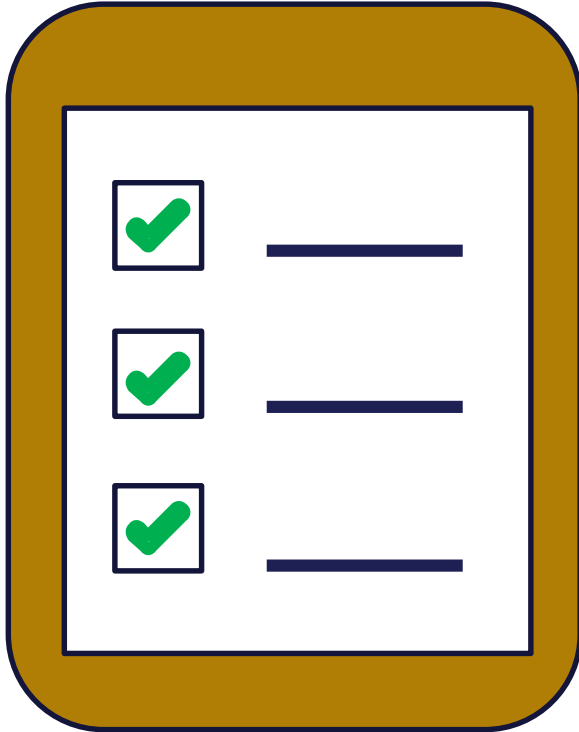
## Over-sampling

- **More observations from minority class.**
- **Can amplify noise**

## Under-sampling

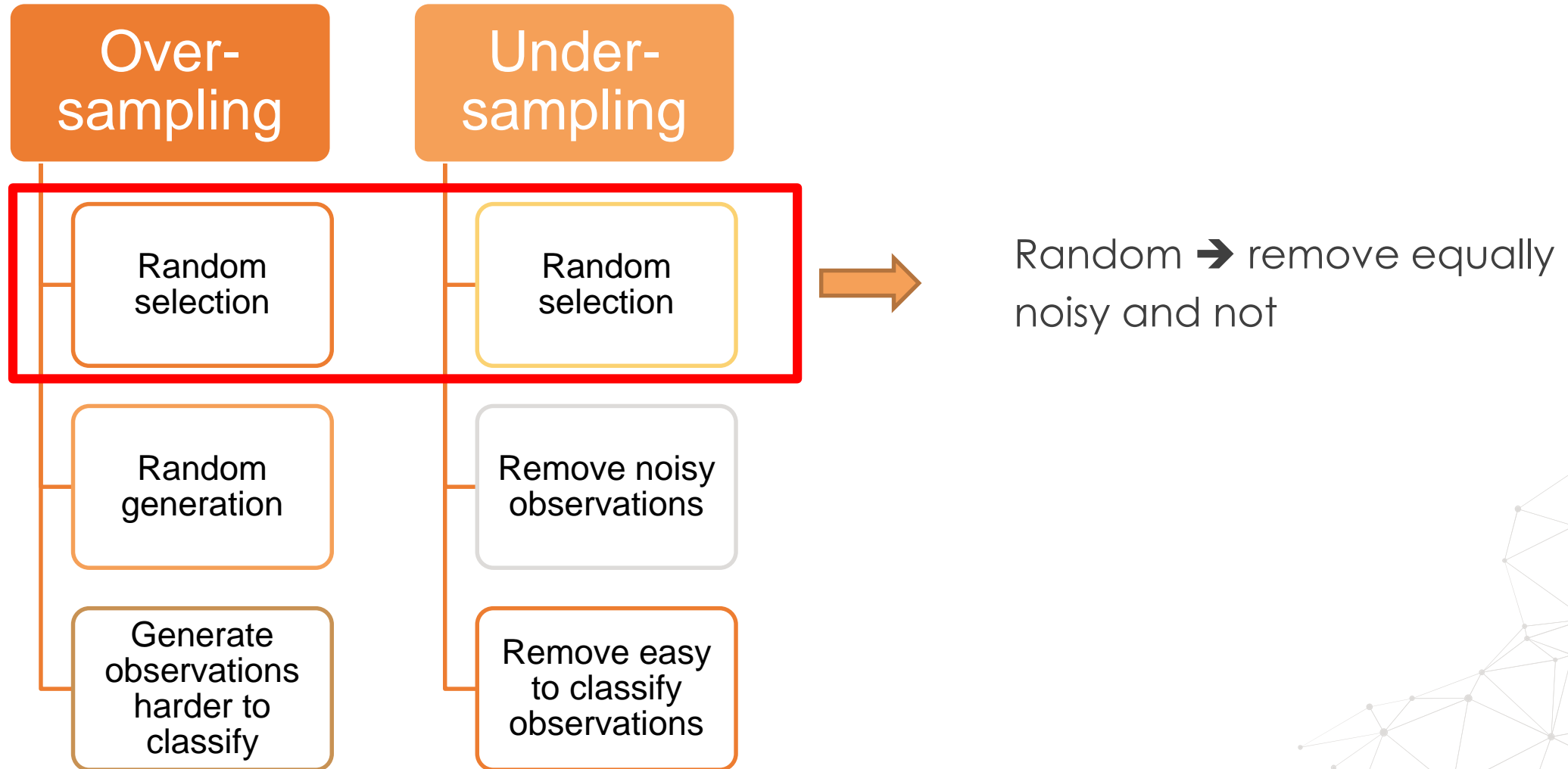
- **Lose important information (from majority class)**
- **Remove noisy observations**

# Over- and under-sampling

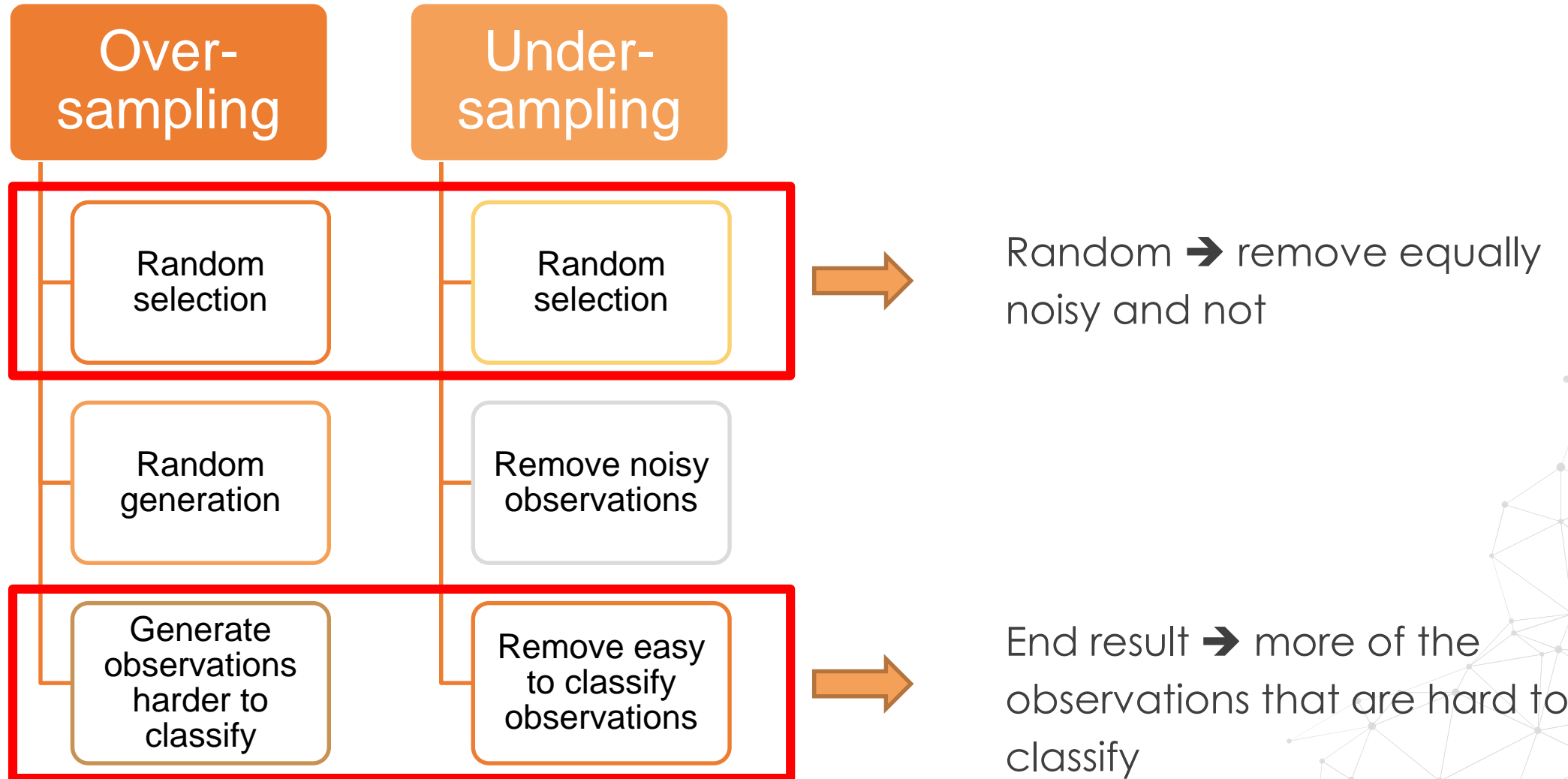


- Retain all observations from the majority
- Increase the number of examples from the minority
- Limit (remove) the impact of noise

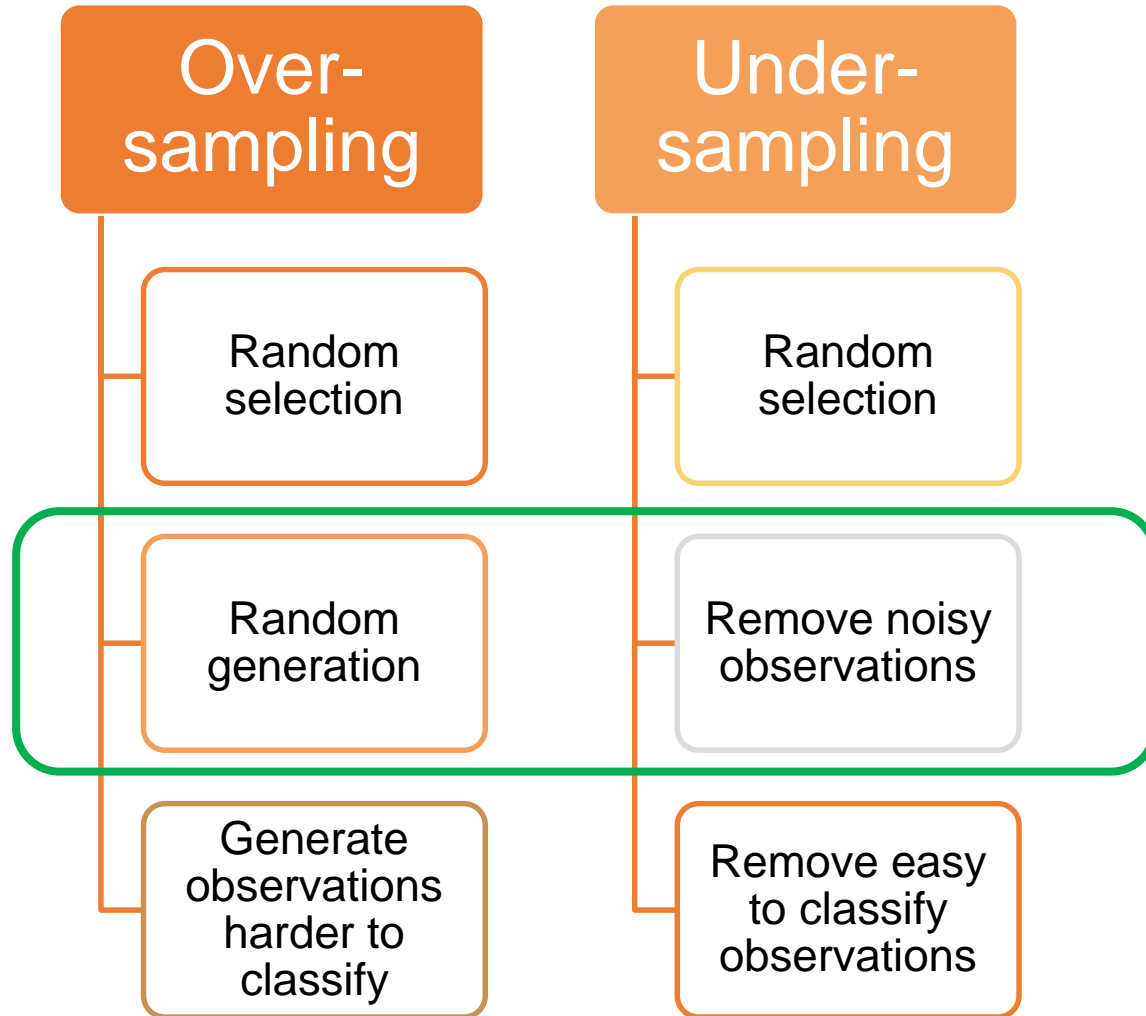
# Over- and under-sampling: which to combine?



# Over- and under-sampling: which to combine?

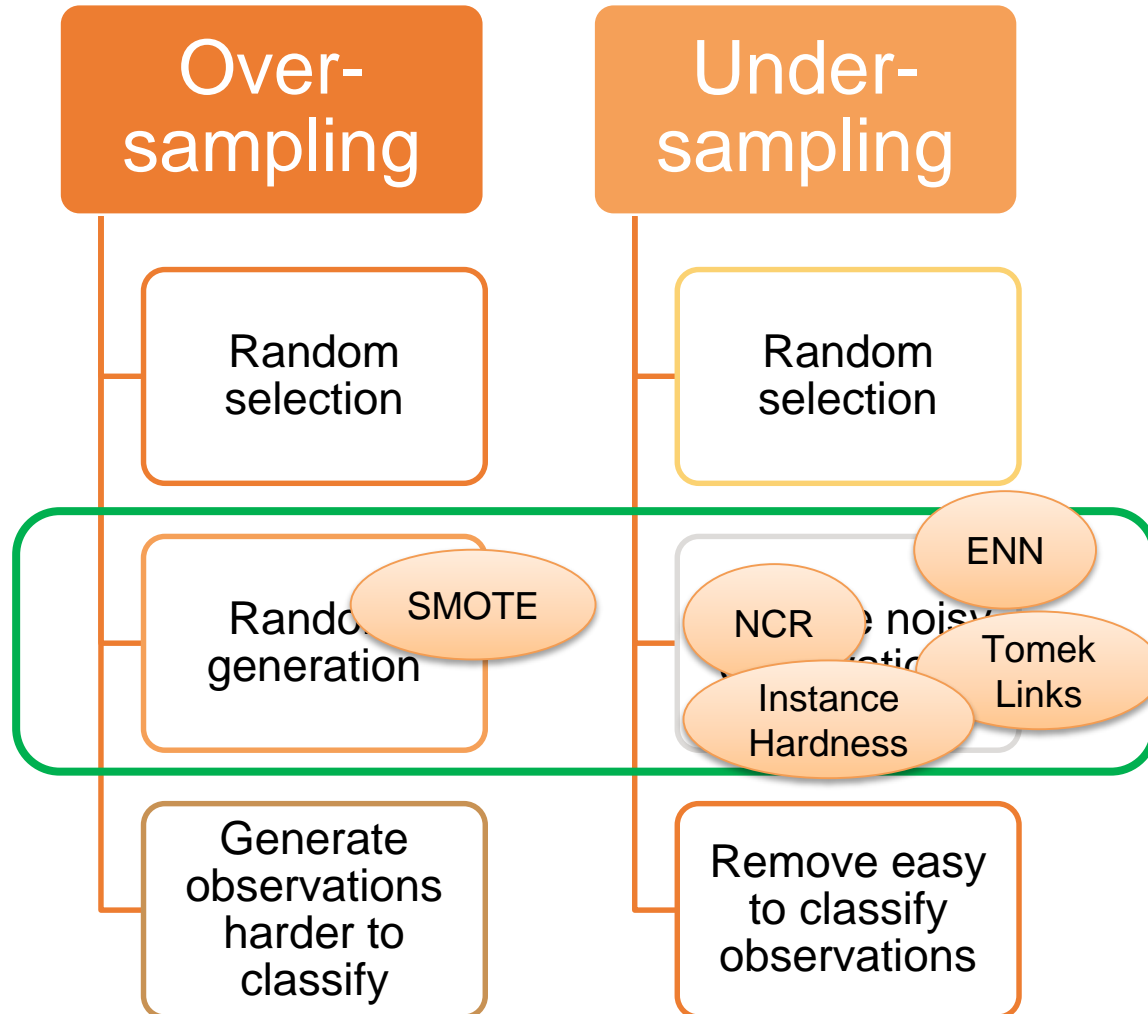


# Over- and under-sampling: which to combine?

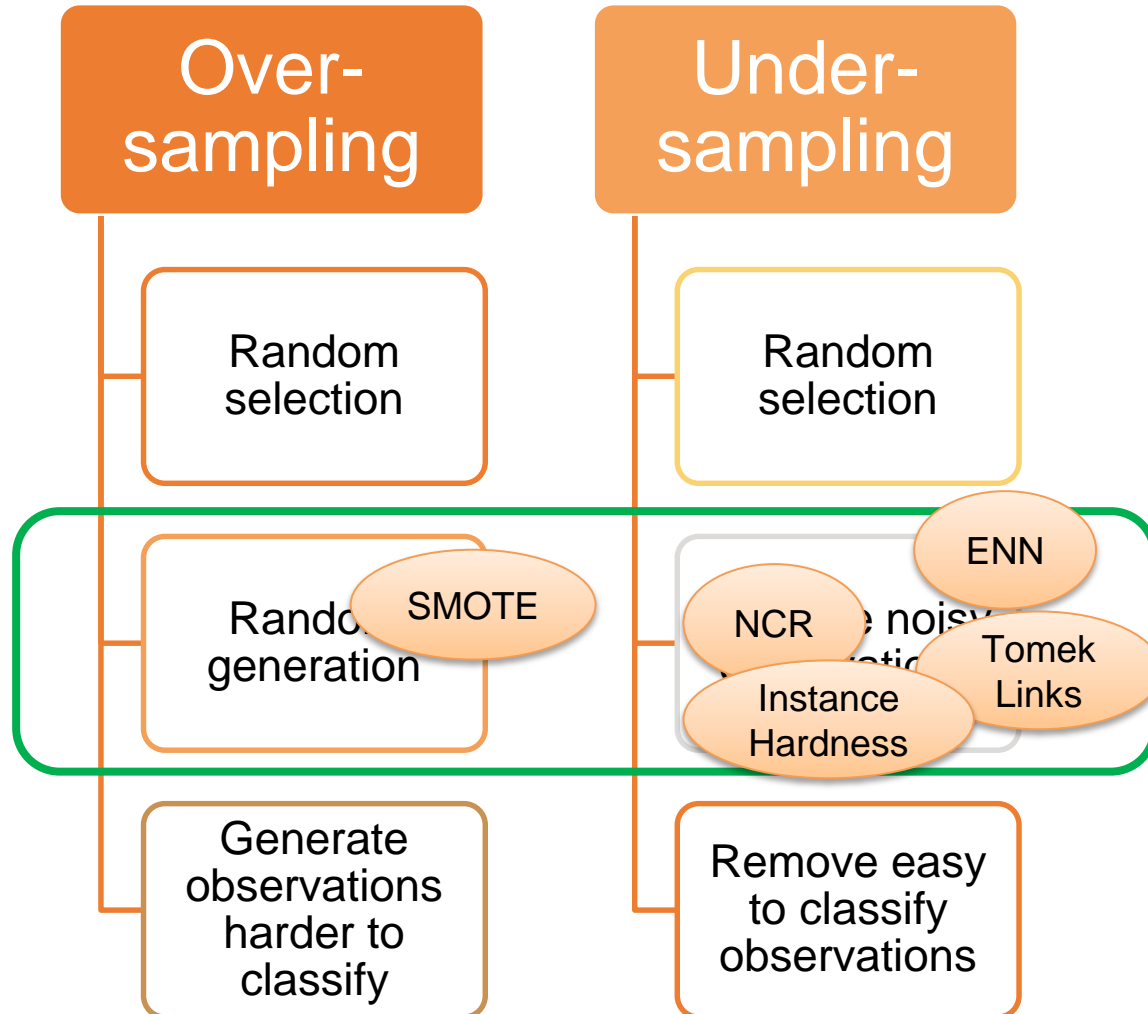




# Over- and under-sampling combination



# Over- and under-sampling combination



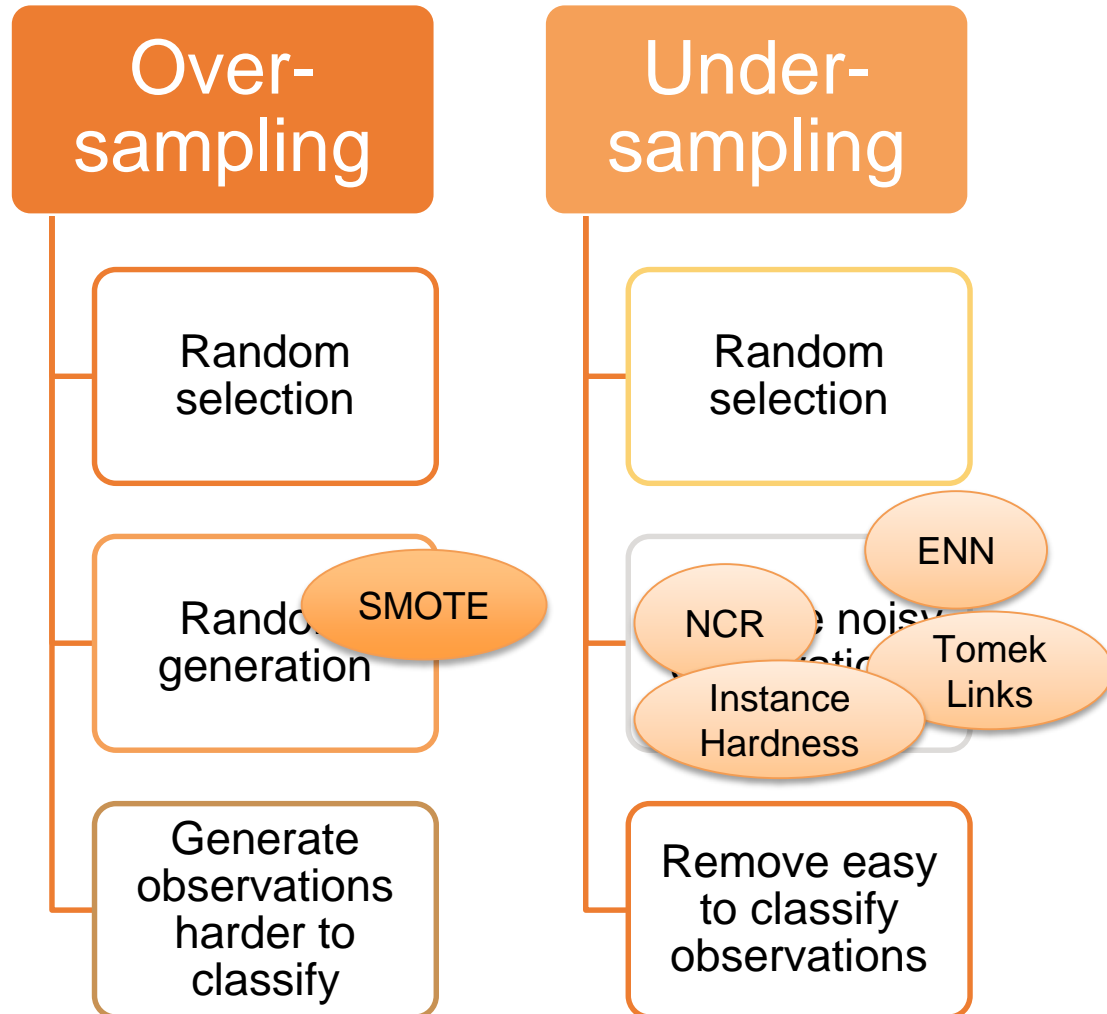
## SMOTE

- Generates observations from all examples, noisy and not
- Could amplify noise

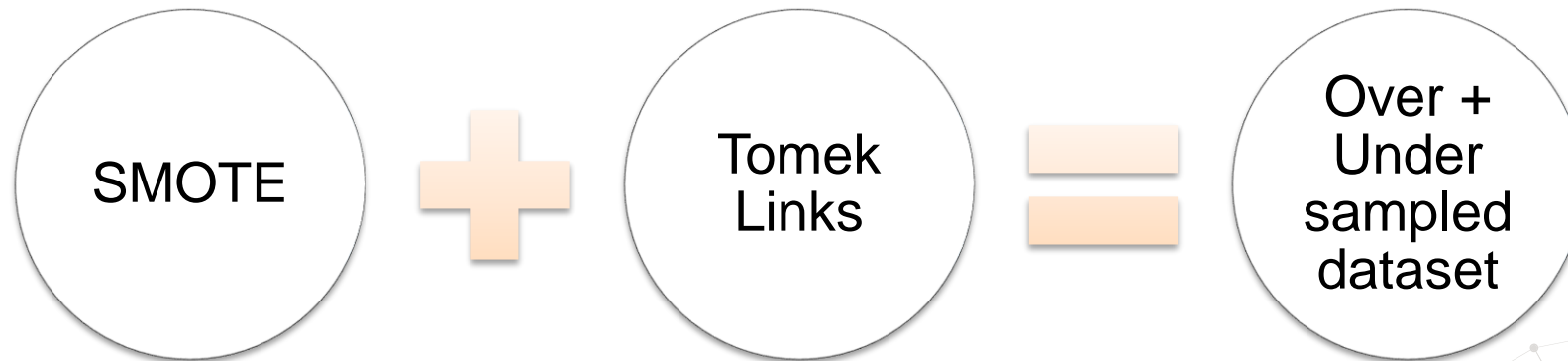
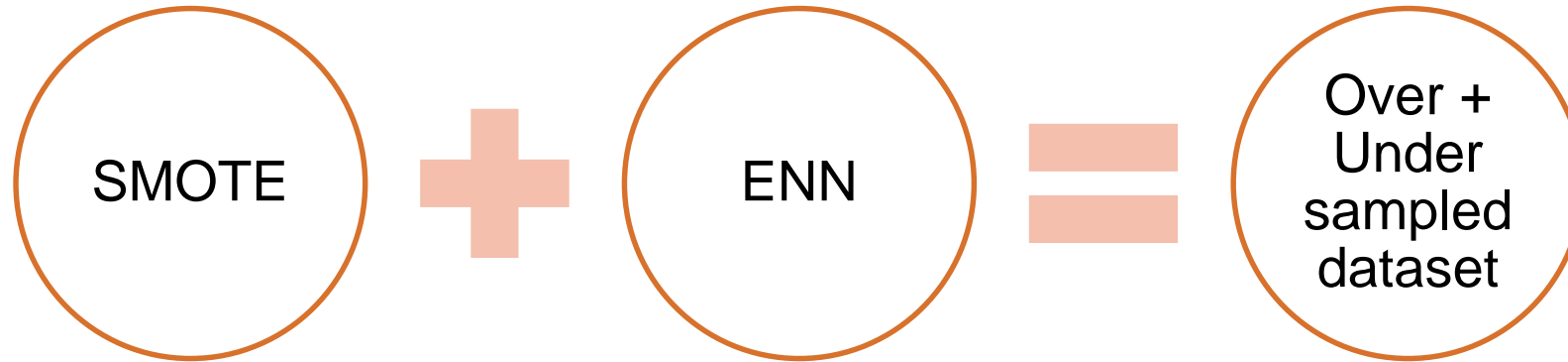
## ENN, Tomek Links

- Remove observations harder to classify → remove noise

# Over- and under-sampling combination



# Over- and under-sampling combination



# THANK YOU

[www.trainindata.com](http://www.trainindata.com)