IT-Academy

СТРАТЕГИИ СЭМПЛИНГА В УСЛОВИЯХ НЕСБАЛАНСИРОВАННОСТИ КЛАССОВ

КУХАЛЬСКИЙ НИКОЛАЙ ГЕННАДЬЕВИЧ

Вопросы занятия

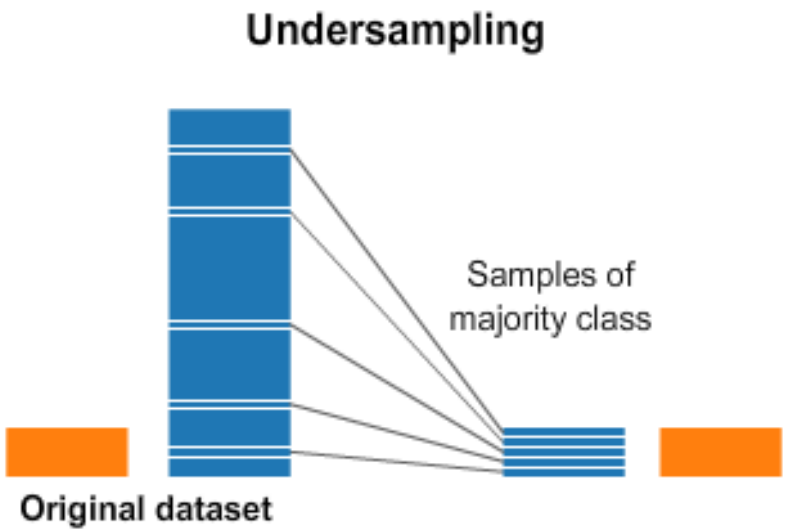
1. Undersampling;

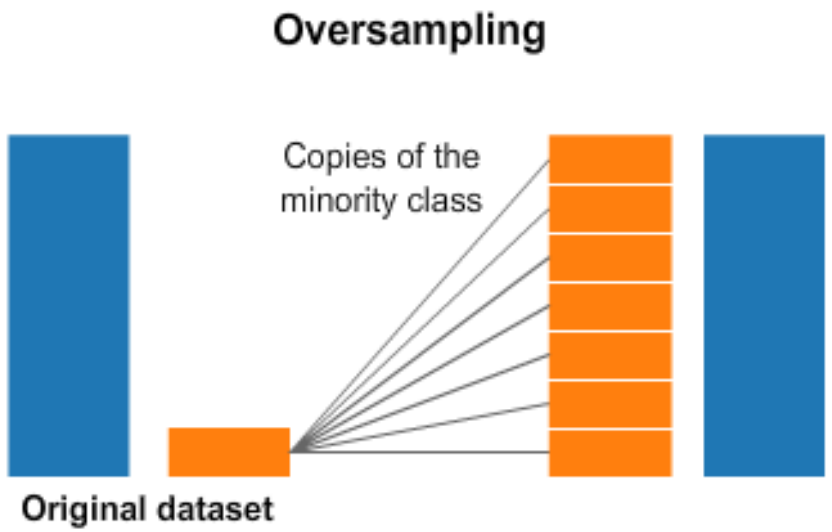
2. Oversampling;

3. Готовые алгоритмы для работы с несбалансированными классами

SAMPLING

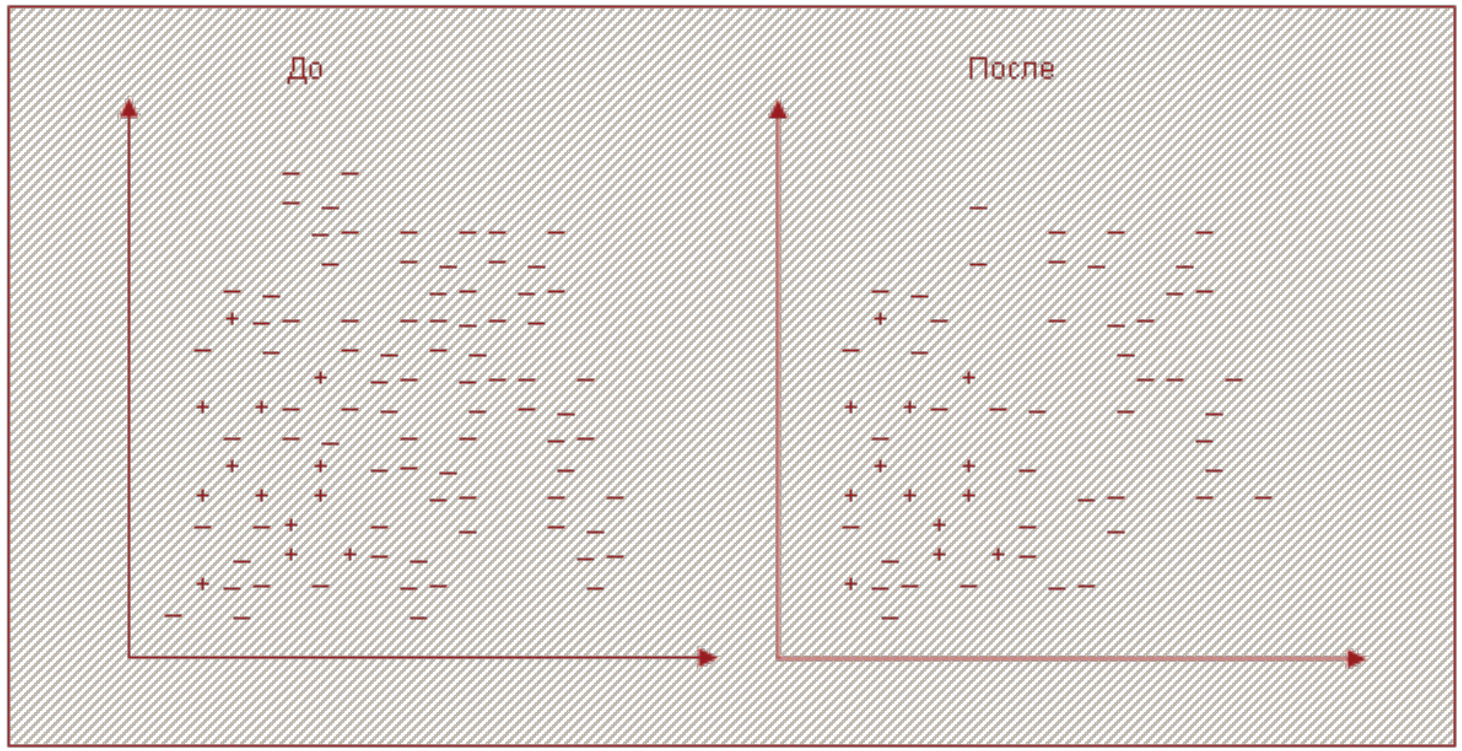
IT-Academy





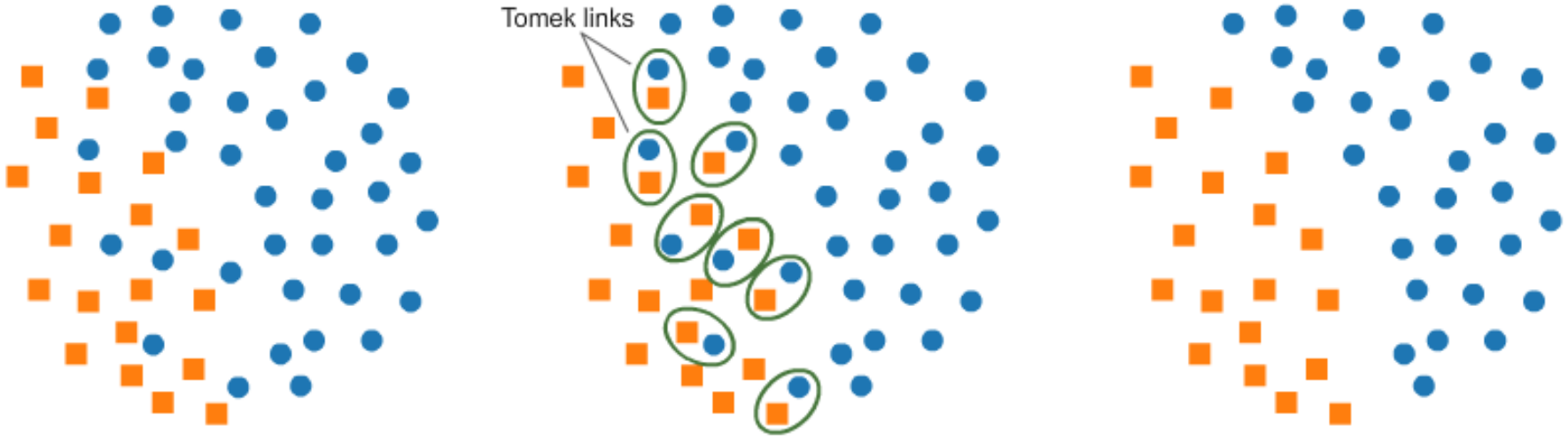
UNDERSAMPLING

Случайное удаление примеров мажоритарного класса (Random Undersampling)



UNDERSAMPLING

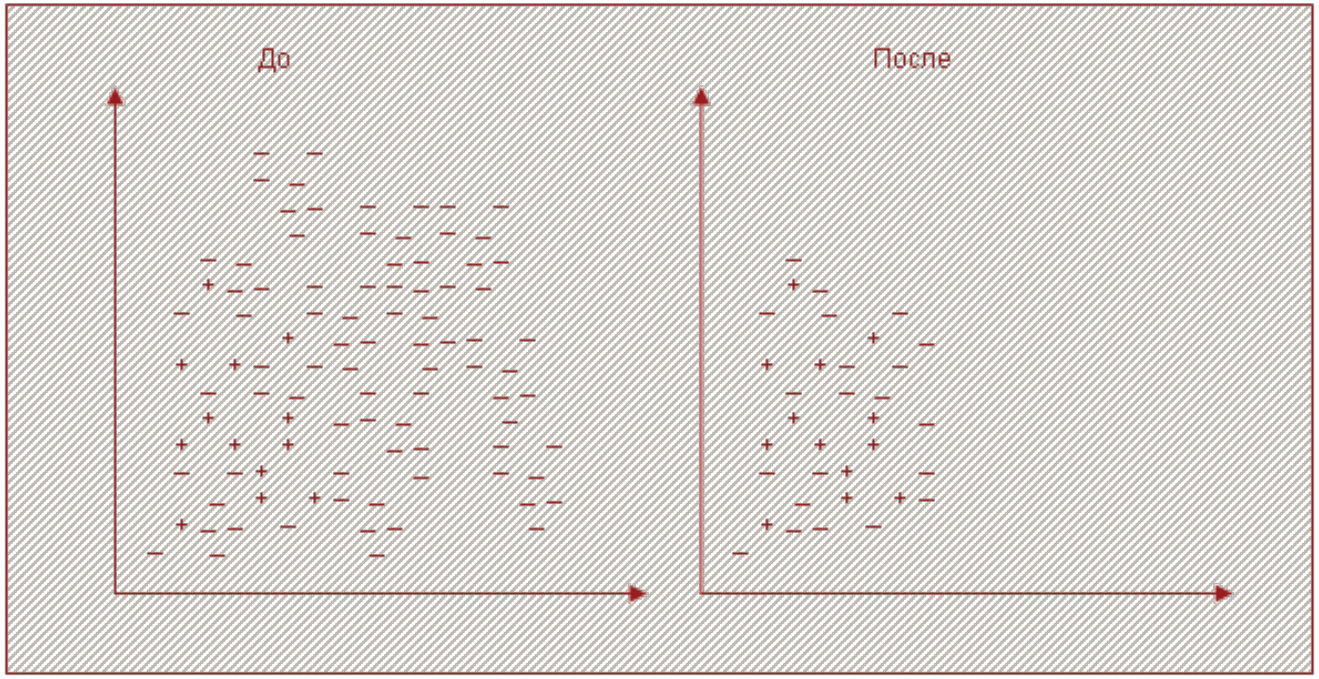
Undersampling Tomek Links



UNDERSAMPLING

Condensed Nearest Neighbor Rule

Классификатор учиться находить отличие между похожими примерами, но принадлежащими к разным классам

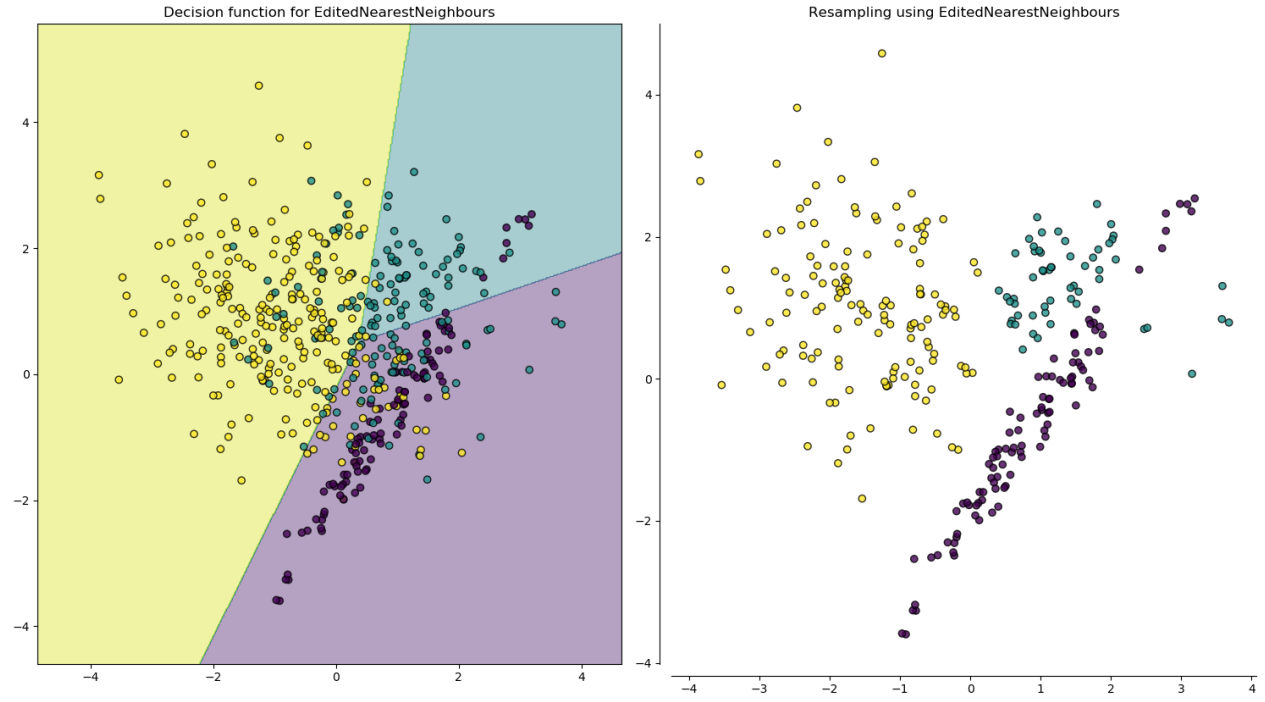


UNDERSAMPLING

IT-Academy

EditedNearestNeighbours

Этот метод произведет очистку, удалив образцы, близкие к границе принятия решения.

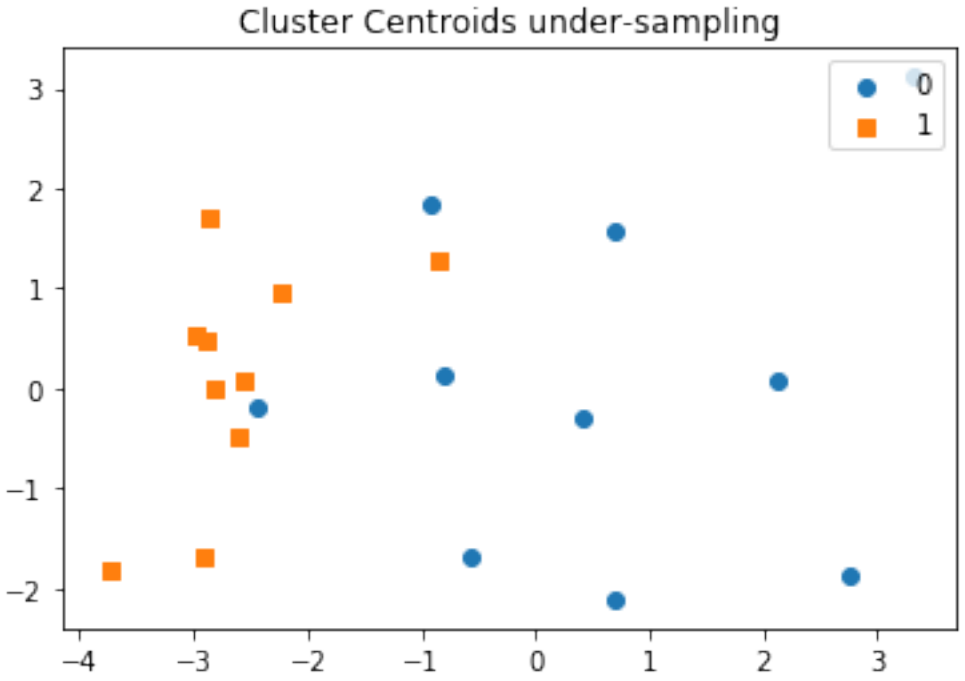


RepeatedEditedNearestNeighbours

Этот метод будет несколько раз повторять алгоритм ENN.

UNDERSAMPLING

Undersampling Cluster Centroids

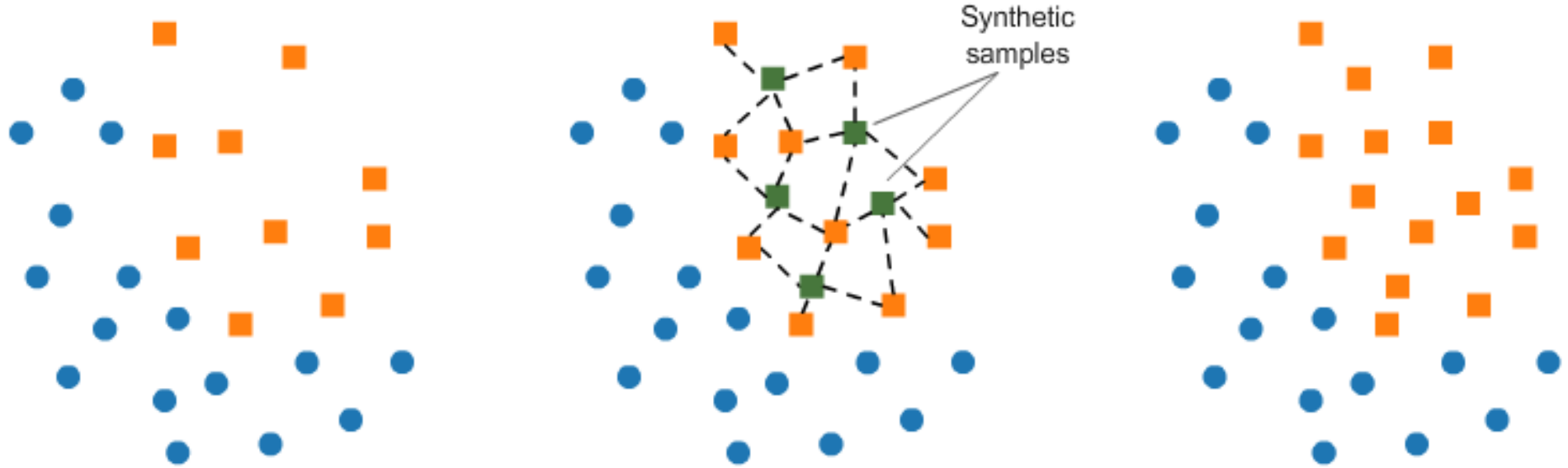


OVERSAMPLING

IT-Academy

SMOTE

Идея генерации некоторого количества искусственных примеров, которые были бы «похожи» на имеющиеся в миноритарном классе, но при этом не дублировали их.

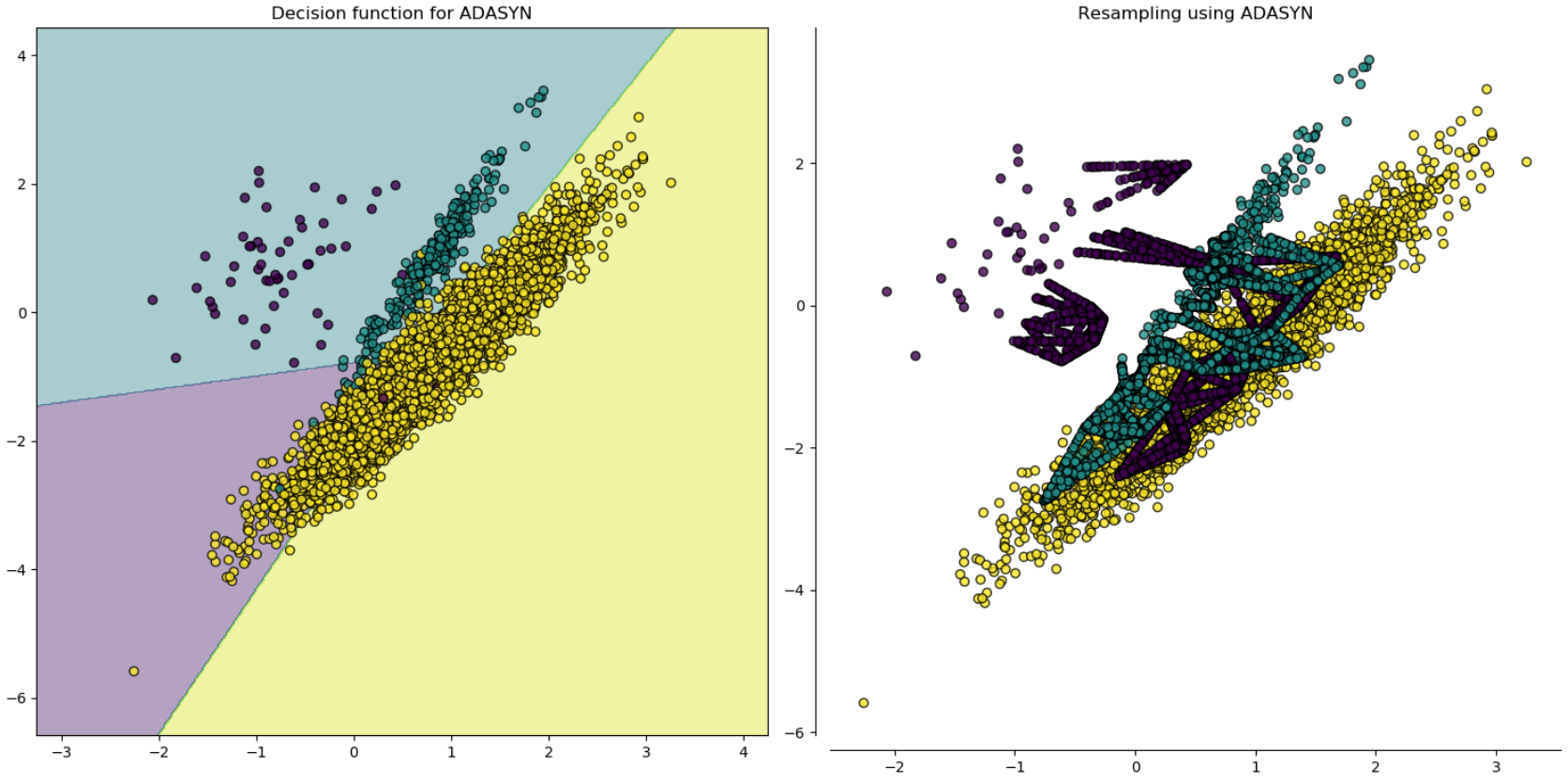


OVERSAMPLING

IT-Academy

ADASYN

Этот метод аналогичен SMOTE, но он генерирует различное количество выборок в зависимости от оценки локального распределения класса, подлежащего передискретизации.

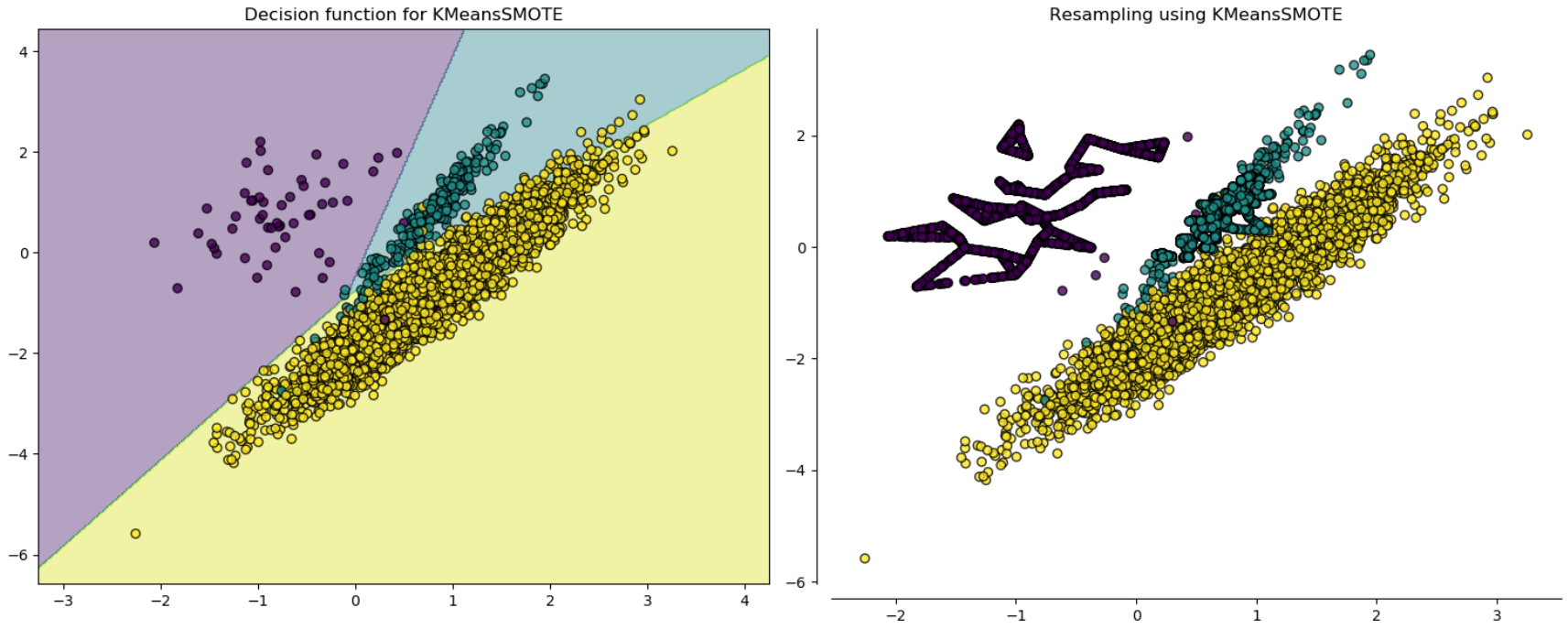


OVERSAMPLING

IT-Academy

KMeansSMOTE

Этот метод использует предварительно KMeans clustering перед применением алгоритма SMOTE.



Combine over- and under-sampling methods

IT-Academy

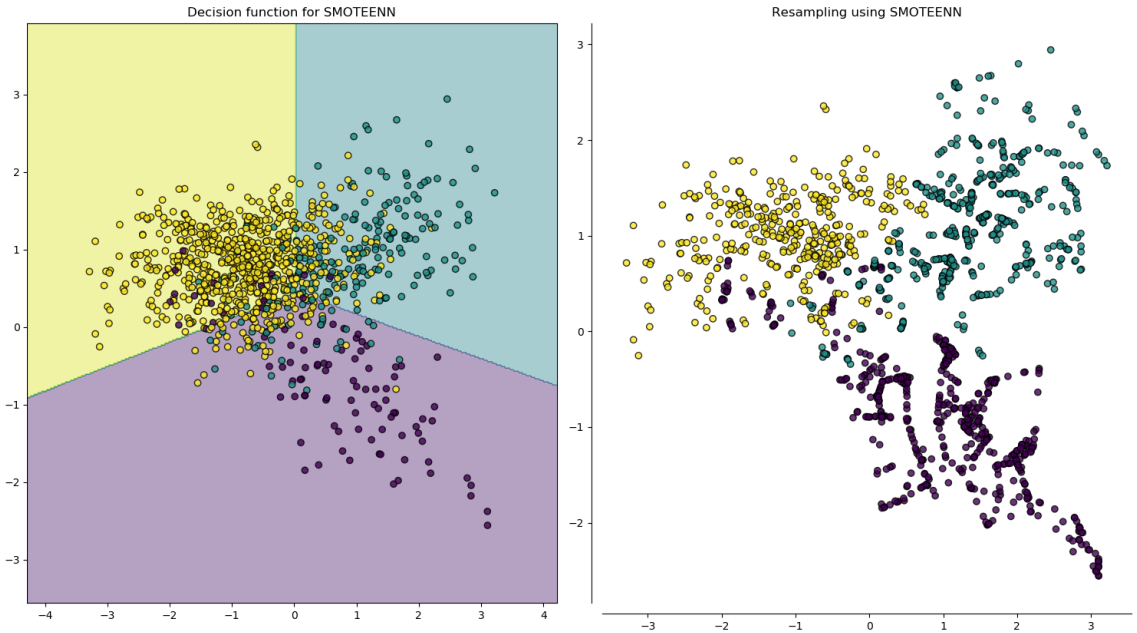
SMOTEENN

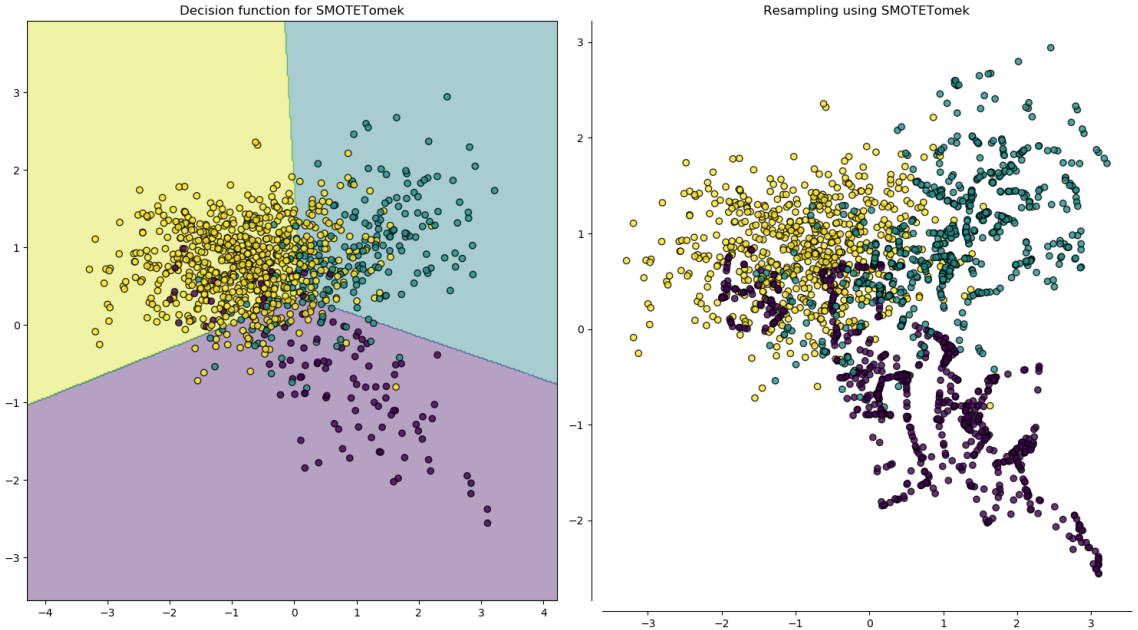
Combine over- and under-sampling using SMOTE and Edited Nearest Neighbours.

SMOTETomek

Over-sampling using SMOTE and cleaning using Tomek links.

Combine over- and under-sampling methods





IT-Academy

Алгоритмы для работы с несбалансированными классами

IT-Academy

imbalanced-learn — это питоновская библиотека для борьбы с проблемами несбалансированных наборов данных.

- **BalancedRandomForestClassifier**
- **BalancedBaggingClassifier**
- **RUSBoostClassifier**
- **EasyEnsembleClassifier**

Помимо общих параметров с базовыми классификаторами имеют параметр работы с несбалансированными классами - ***sampling_strategy.***

IT-Academy

ПРАКТИКА

winequality-red.csv

IT-Academy

СТРАТЕГИИ СЭМПЛИНГА В УСЛОВИЯХ НЕСБАЛАНСИРОВАННОСТИ КЛАССОВ

КУХАЛЬСКИЙ НИКОЛАЙ ГЕННАДЬЕВИЧ