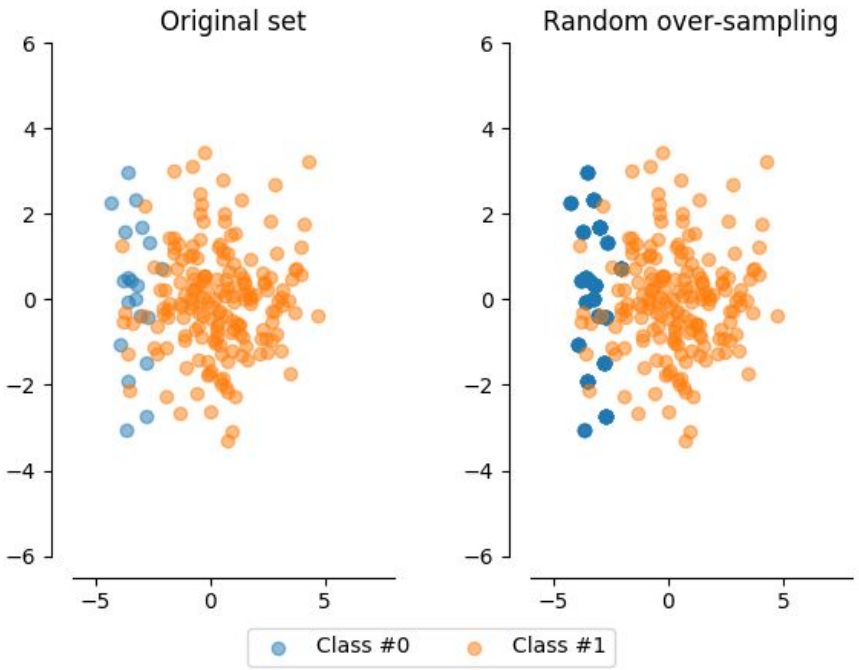
TESTES DE MÉTODOS DE *OVERSAMPLING* PARA *DATASETS* DESBALANCEADOS

Este documento contém resultados de testes de aplicação de métodos de *oversampling* utilizando a biblioteca **imbalanced-learn** (https://github.com/scikit-learn-contrib/imbalanced-learn) na linguagem Python em *dataset* desbalanceado com 2 classes (#0 e #1) gerado pela própria plataforma scikit-learn através sklearn.datasets.make_classification.

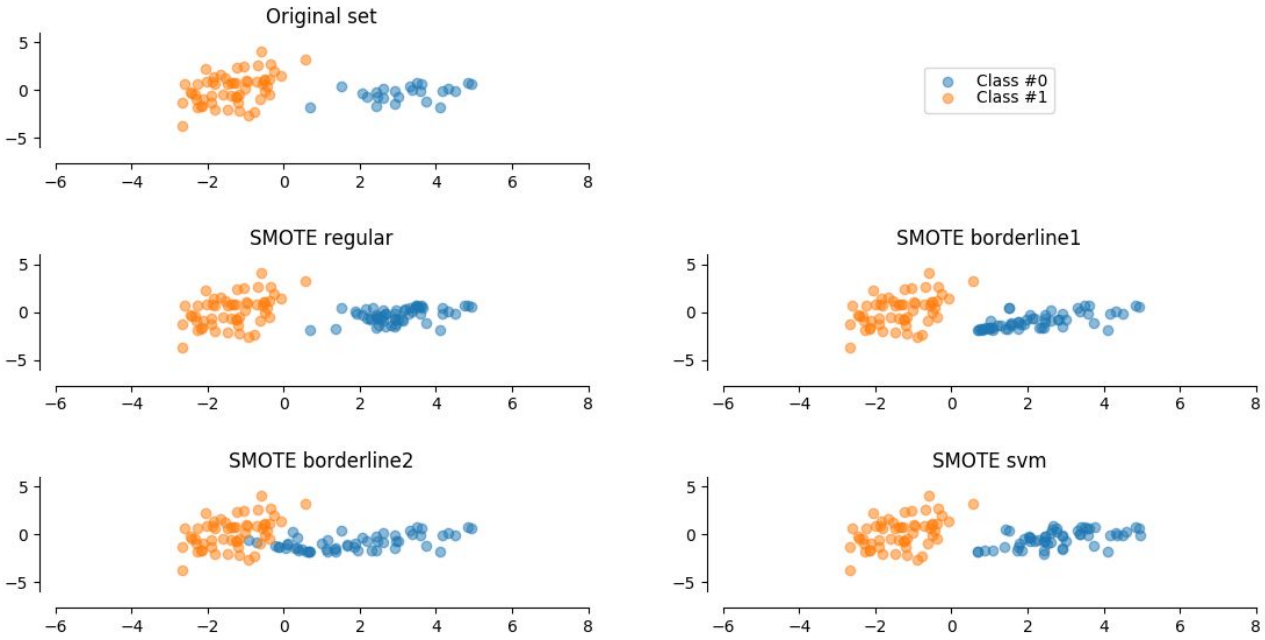
Os métodos testados para *oversampling* foram 6, sendo 4 deles variações do algoritmo SMOTE. São eles:

1. **Random minority over-sampling with replacement**
   - http://contrib.scikit-learn.org/imbalanced-learn/generated/imblearn.over_sampling.RandomOverSampler.html

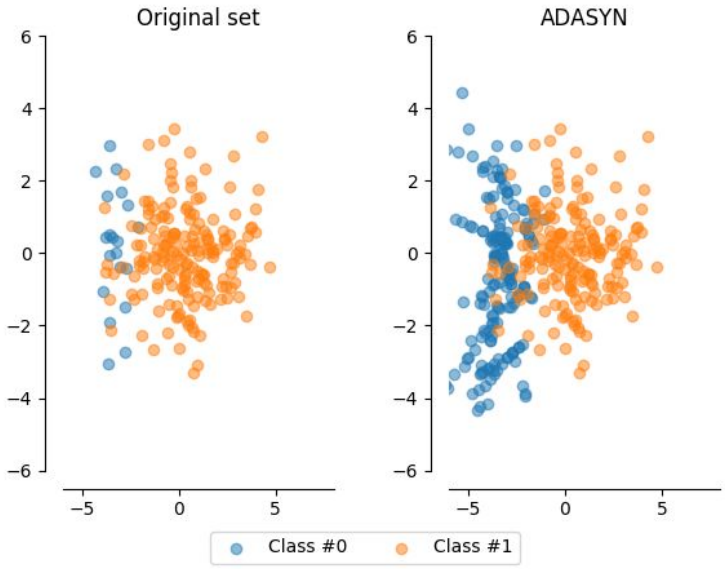


2. **SMOTE - Synthetic Minority Over-sampling Technique**
3. **bSMOTE(1 & 2) - Borderline SMOTE of types 1 and 2**
4. **SVM SMOTE - Support Vectors SMOTE**

- http://contrib.scikit-learn.org/imbalanced-learn/generated/imblearn.over_sampling.SMOTE.html

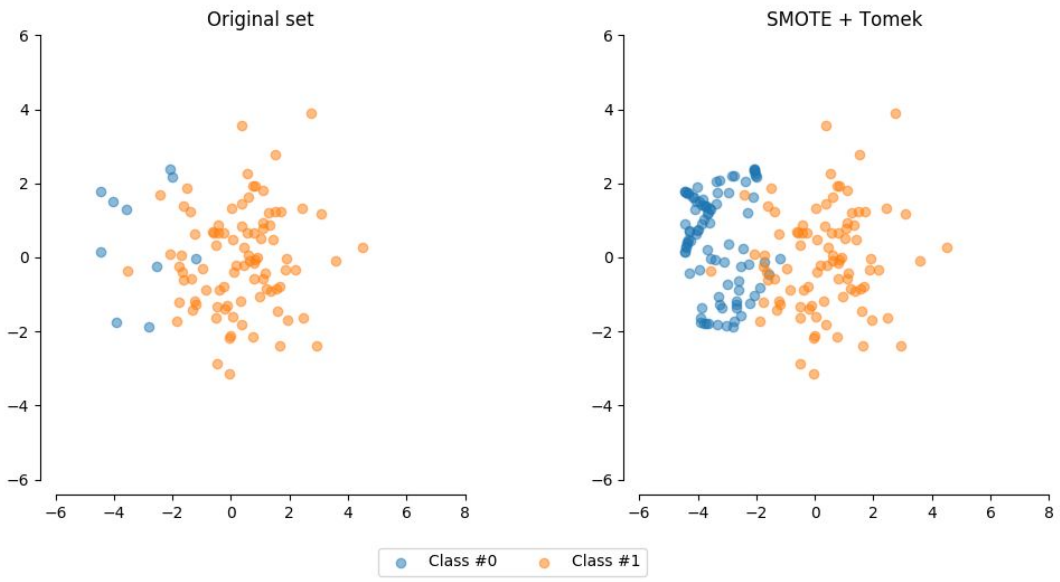


5. **ADASYN - Adaptive synthetic sampling approach for imbalanced learning**
   - http://contrib.scikit-learn.org/imbalanced-learn/generated/imblearn.over_sampling.ADASYN.html

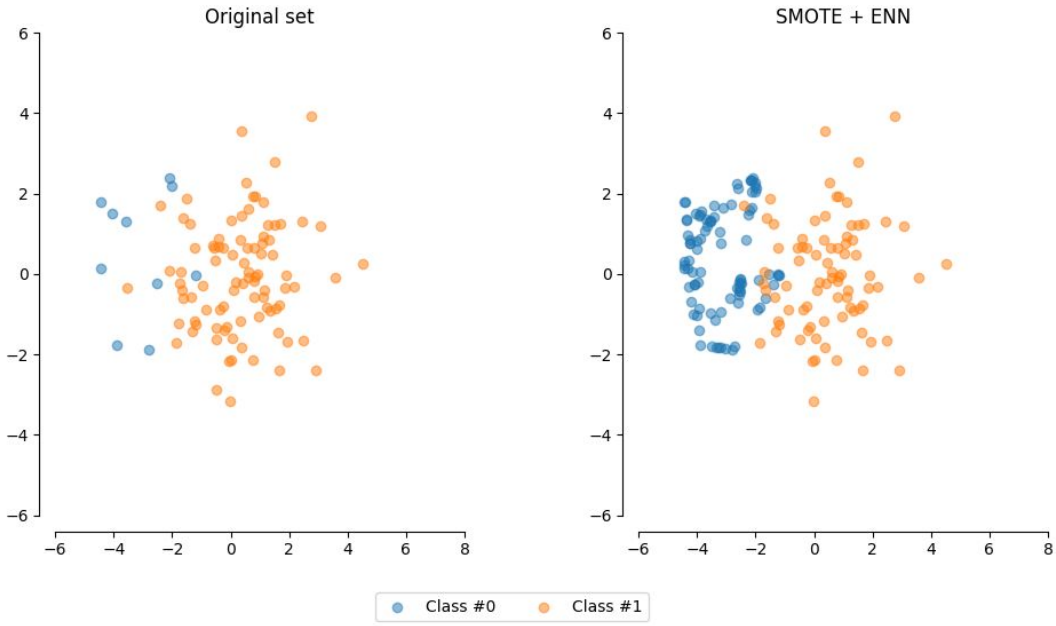


Os métodos testados para *oversampling* com *undersampling* foram 2, com ambos os métodos utilizando SMOTE:

1. **SMOTE + Tomek links**
    - http://contrib.scikit-learn.org/imbalanced-learn/generated/imblearn.combine.SMOTETomek.html



2. **SMOTE + ENN**
    - http://contrib.scikit-learn.org/imbalanced-learn/generated/imblearn.combine.SMOTEENN.html


Original set
SMOTE + ENN
Class #0
Class #1