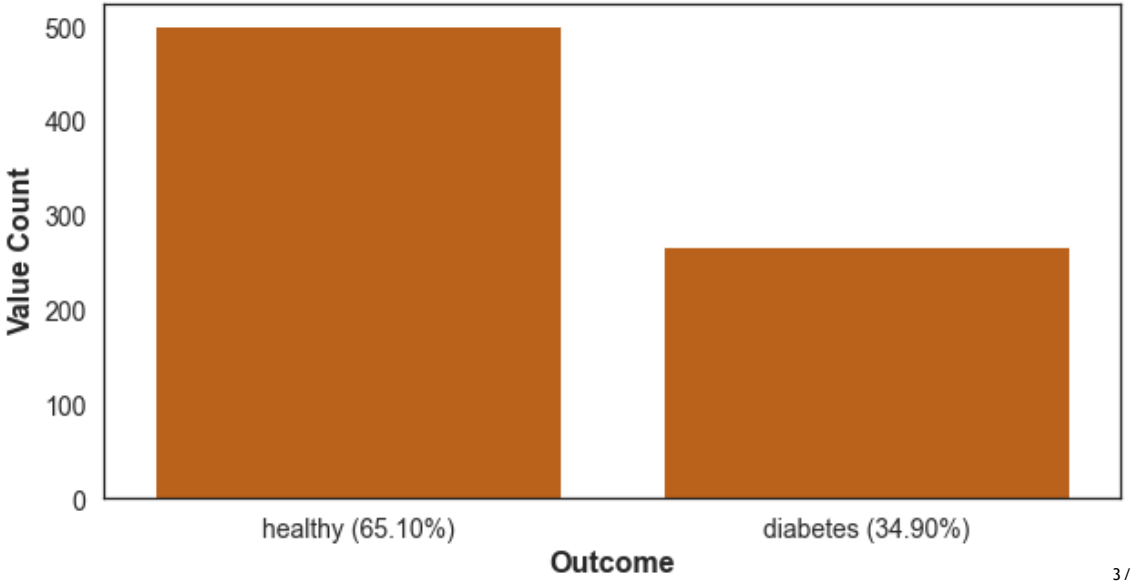
SMOTE with Active Learning

Master IASD
Raul Sena Rojas

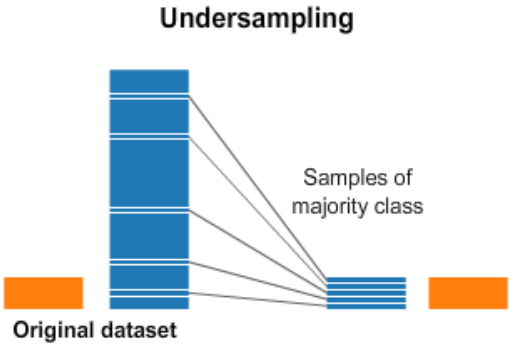
Contents

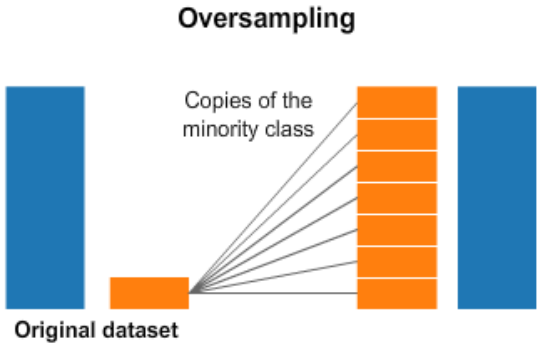
Class Imbalance Problem in Medical Datasets - PIMA



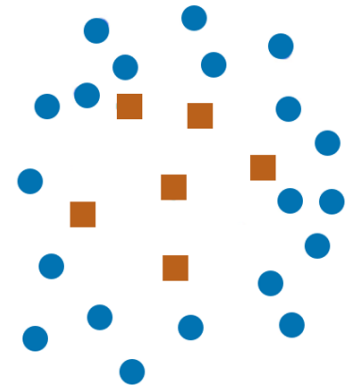
Classical sampling techniques to handle Class Imbalance

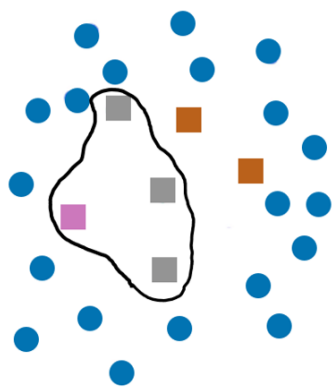
Possible Solutions: “Sampling Techniques”



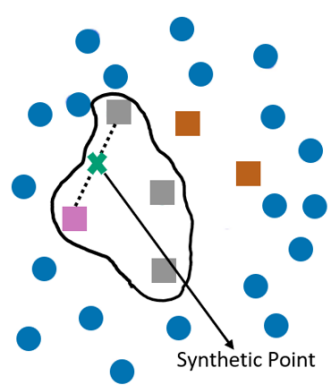


SMOTE (Synthetic Minority Oversampling Technique)



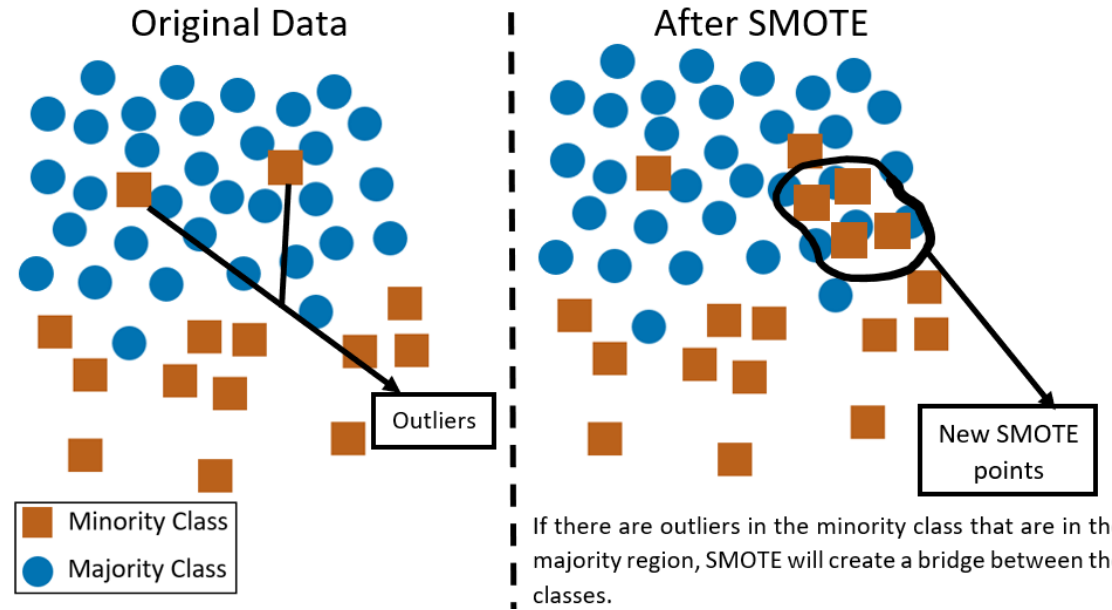


Randomly select a minority class sample (pink point) and its k nearest neighbors (gray points, k=3)



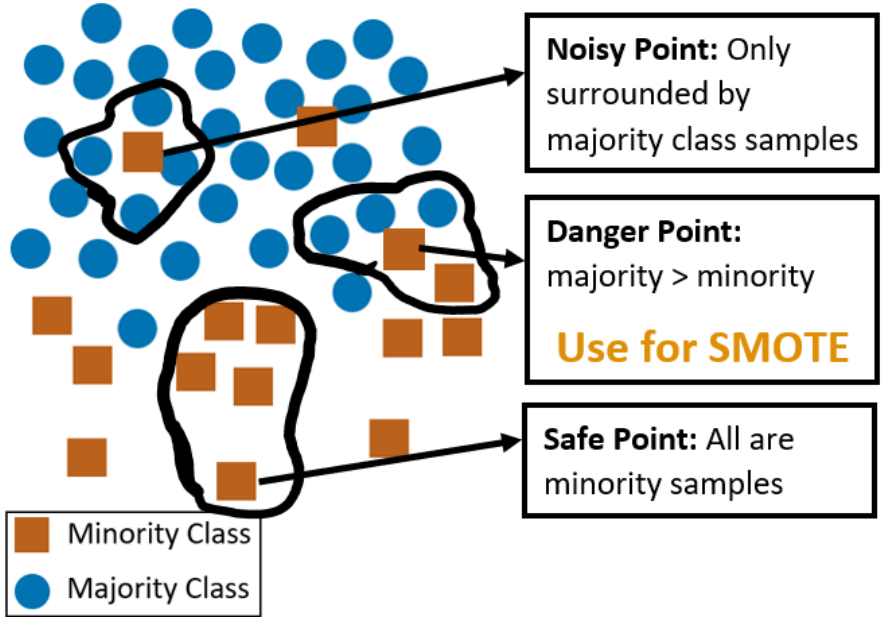
Randomly select a k-neighbor and generate a synthetic point (green x) along the dotted line.

SMOTE inconveniences



If there are outliers in the minority class that are in the majority region, SMOTE will create a bridge between the classes.

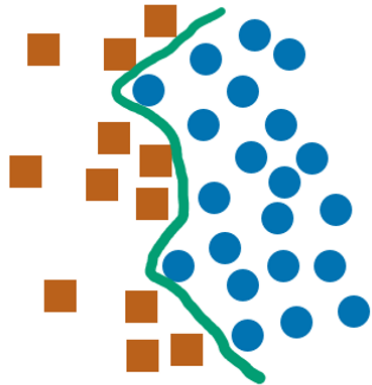
Borderline SMOTE



SMOTE with Active Learning

“The new proposed algorithm”

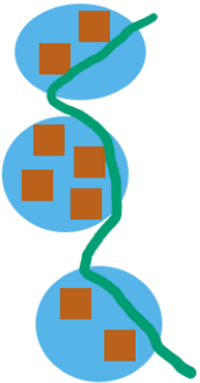
Uncertainty Sampling Phase



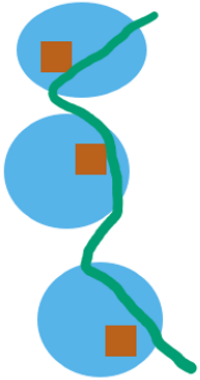
Train a Machine Learning classifier and obtain a decision boundary.

Sample from the minority class the items where the model is most uncertain.

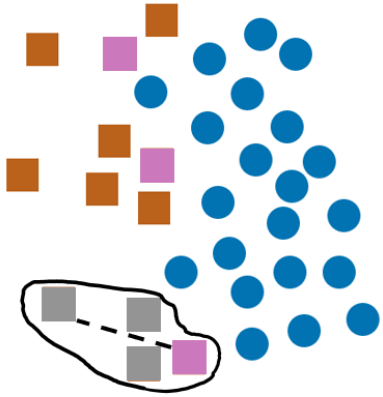
Diversity Sampling Phase



Cluster the items with K-means.

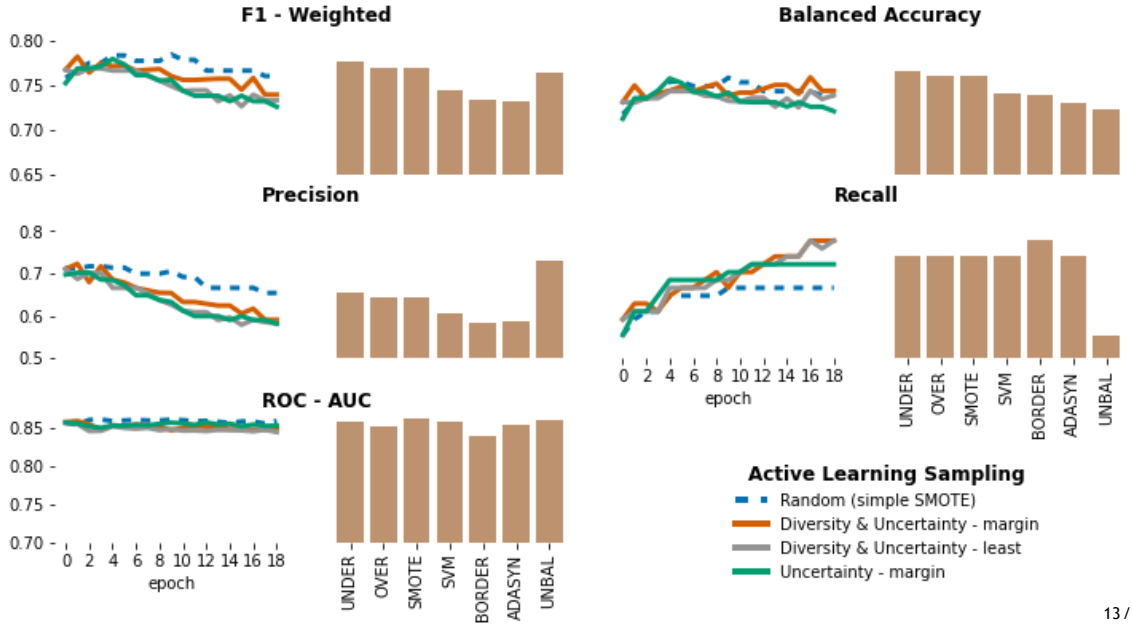


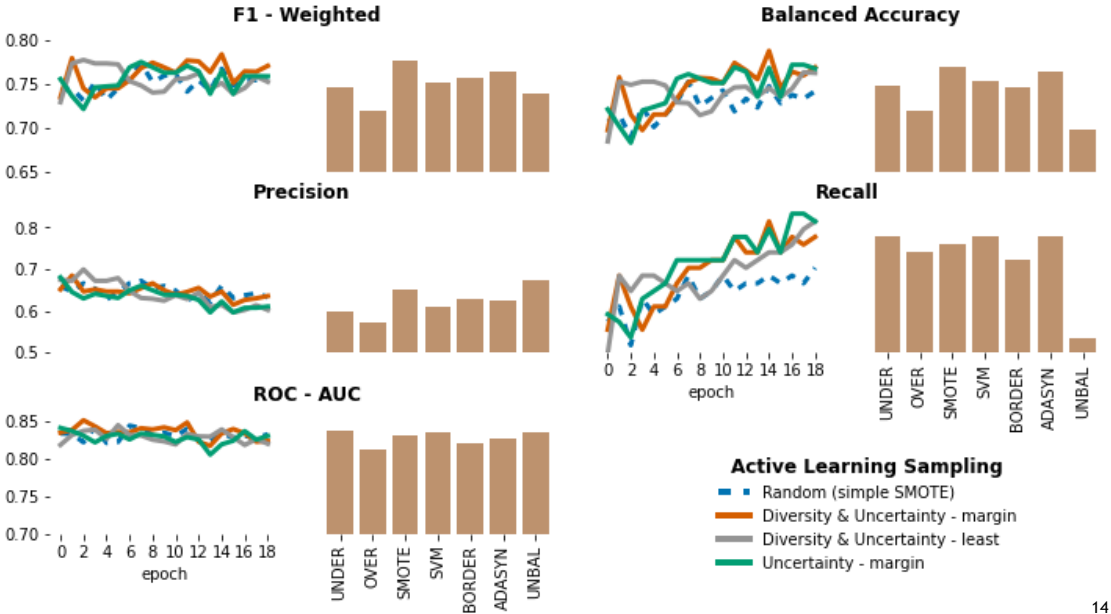
From each cluster, choose N samples randomly.

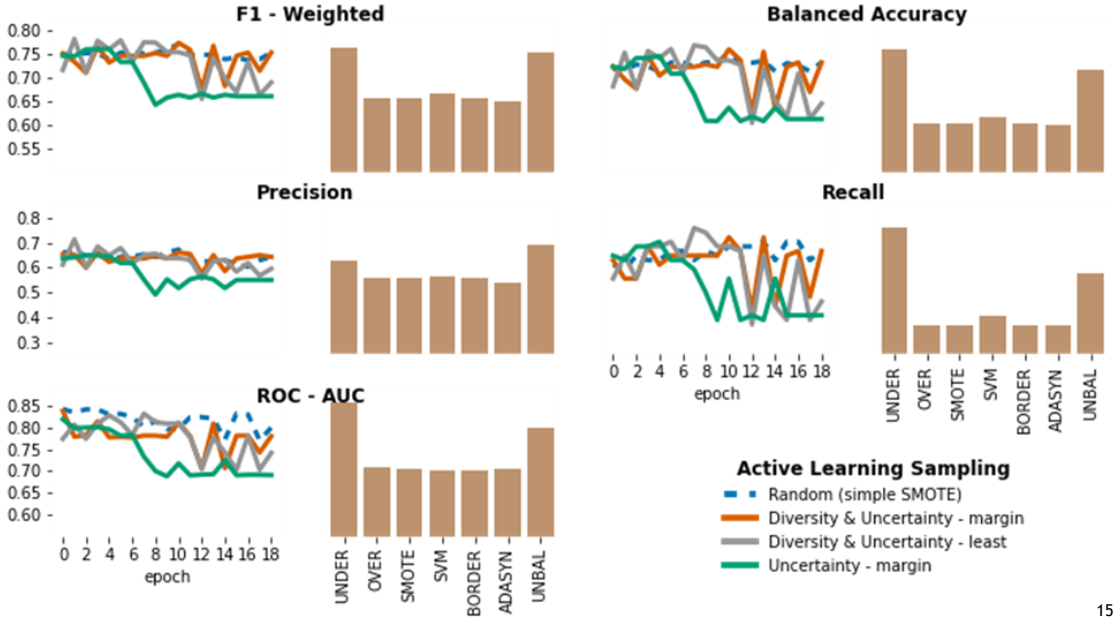


Use those points as the initial ones in classical SMOTE.

Results SMOTE with Active Learning


Logistic Regression - PIMA - Test
F1 - Weighted
Balanced Accuracy
Precision
Recall
ROC - AUC
epoch
UNDER
OVER
SMOTE
SVM
BORDER
ADASYN
UNBAL
Active Learning Sampling
Random (simple SMOTE)
Diversity & Uncertainty - margin
Diversity & Uncertainty - least
Uncertainty - margin


Gradient Boost - PIMA - Test
F1 - Weighted
Balanced Accuracy
Precision
Recall
ROC - AUC
epoch
UNDER
OVER
SMOTE
SVM
BORDER
ADASYN
UNBAL
Active Learning Sampling
Random (simple SMOTE)
Diversity & Uncertainty - margin
Diversity & Uncertainty - least
Uncertainty - margin


SVC - PIMA - Test
F1 - Weighted
Balanced Accuracy
Precision
Recall
ROC - AUC
epoch
UNDER
OVER
SMOTE
SVM
BORDER
ADASYN
UNBAL
Active Learning Sampling
Random (simple SMOTE)
Diversity & Uncertainty - margin
Diversity & Uncertainty - least
Uncertainty - margin


Ada Boost - PIMA - Test
F1 - Weighted
Balanced Accuracy
Precision
Recall
ROC - AUC
epoch
UNDER
OVER
SMOTE
SVM
BORDER
ADASYN
UNBAL
Active Learning Sampling
Random (simple SMOTE)
Diversity & Uncertainty - margin
Diversity & Uncertainty - least
Uncertainty - margin


Random Forest - PIMA - Test
F1 - Weighted
Balanced Accuracy
Precision
Recall
ROC - AUC
epoch
UNDER
OVER
SMOTE
SVM
BORDER
ADASYN
UNBAL
Active Learning Sampling
Random (simple SMOTE)
Diversity & Uncertainty - margin
Diversity & Uncertainty - least
Uncertainty - margin