

ADASYN - Oversampling

```
def adasyn( data, y, k = 5, samp_method = "balance", drop_na_col = True,  
drop_na_row = True, rel_thres = 0.5, rel_method = "auto", rel_xtrm_type = "both",  
rel_coef = 1.5, rel_ctrl_pts_rg = None):
```

ADASYN represents the Adaptive Synthetic algorithm and oversamples the minority class depending on an estimate of the local distribution of the class.

Function designed to help solve the problem of imbalanced data for regression.

Parameters:

main arguments / inputs:

data: pandas dataframe, the training set.

y: string, response variable y by name. It should be a header name found in the dataframe data.

k: int (positive), default = 5, number of neighbors for over-sampling.

samp_method: {'balance', 'extreme'}, default = 'balance', specified method to determine over sampling percentage.

drop_na_col: bool, default = 'True', if 'True', auto drop columns with NaN's.

drop_na_row: bool, default = 'True', if 'True', auto drop rows with NaN's.

phi relevance function arguments / inputs:

rel_thres: float, positive real number, default = 0.5, define the relevance threshold considered rare in phi relevance function.

rel_method: {'auto', 'manual'}, default = 'auto', the relevance method in phi relevance function.

rel_xtrm_type: {'low', 'high', 'both'}, default = 'both', distribution focus on high, low or both.

rel_coef: float, positive real number, default = 1.5, coefficient for box plot in phi relevance function to consider rare.

rel_ctrl_pts_rg: 2d array, default = None, when rel_method = 'manual', it inputs for "manual" rel method.

References:

He, H., Bai, Y., Garcia, E. A., & Li, S. (2008, June). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In 2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence) (pp. 1322-1328). IEEE.

<https://www.ele.uri.edu/faculty/he/PDFfiles/adasyn.pdf>.

Branco, P., Torgo, L., Ribeiro, R. (2017). SMOGN: A Pre-Processing Approach for Imbalanced Regression. Proceedings of Machine Learning Research, 74:36-50.

<http://proceedings.mlr.press/v74/branco17a/branco17a.pdf>.

Kunz, N. (2019). SMOGN: Synthetic Minority Over-Sampling for Regression with Gaussian Noise (Version 0.1.0). Python Package

Index. <https://pypi.org/project/smogn>.