



Borderline SMOTE

Borderline SMOTE

Is an extension of SMOTE which creates synthetic examples **only** from observations in the minority class **closer to the boundary** with the majority class or classes.

Presents 2 variants, borderline 1 and 2.

Borderline SMOTE

1. Fits KNN with all the dataset (majority + minority observations)
2. Finds and ignores observations from the minority class which K neighbours belong to the majority class → noise and thus irrelevant
3. Finds and ignores observations from the minority class if most of their neighbours are from the minority class → safe and easy to classify
4. **Selects the observations of the minority class if most of their neighbours are from the majority class**

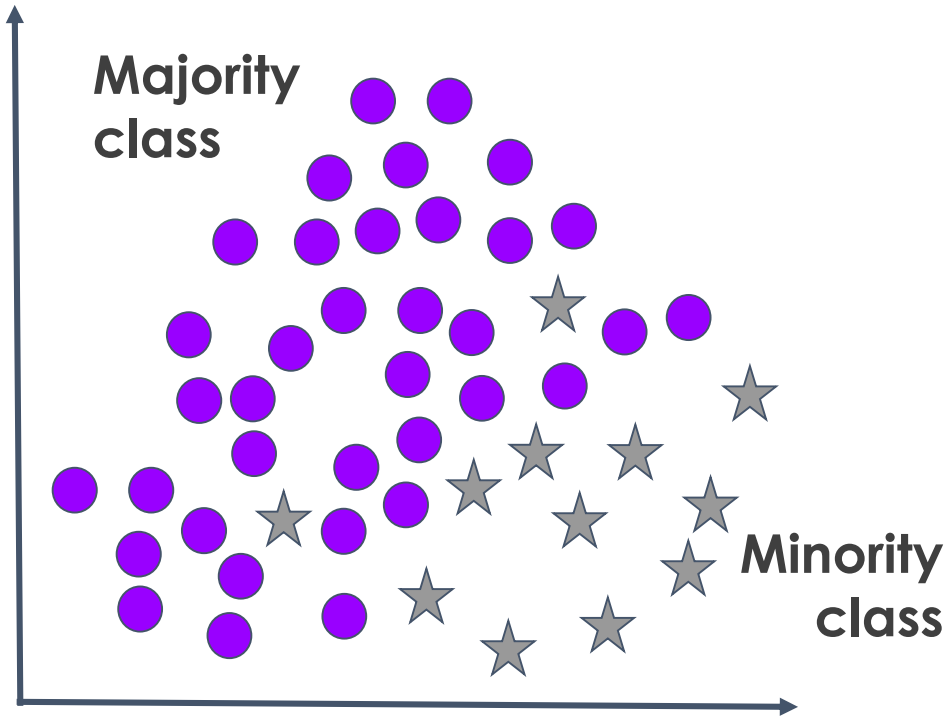
Borderline SMOTE – variant 1

5. Fits KNN to minority class examples (all minority examples)
6. **Variant 1:** Interpolates synthetic samples as SMOTE, between the observations in the DANGER group and its neighbours from the minority class (DANGER group or not)

Borderline SMOTE – variant 2

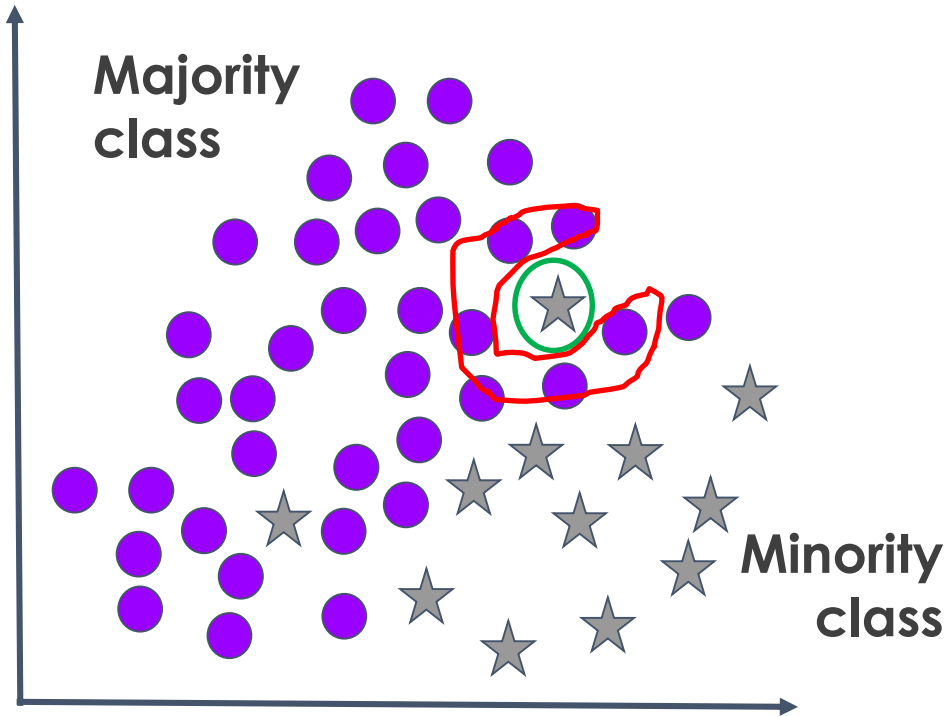
5. Fits KNN to minority class examples (all minority examples)
 6. Interpolates synthetic samples as SMOTE, between the observations in the DANGER group and its neighbours from the minority class (DANGER group or not)
- +
7. Interpolates synthetic examples between the observations in the DANGER group and its neighbours in the majority class, but closer to the DANGER group

Borderline SMOTE: visually



Fits KNN to entire dataset

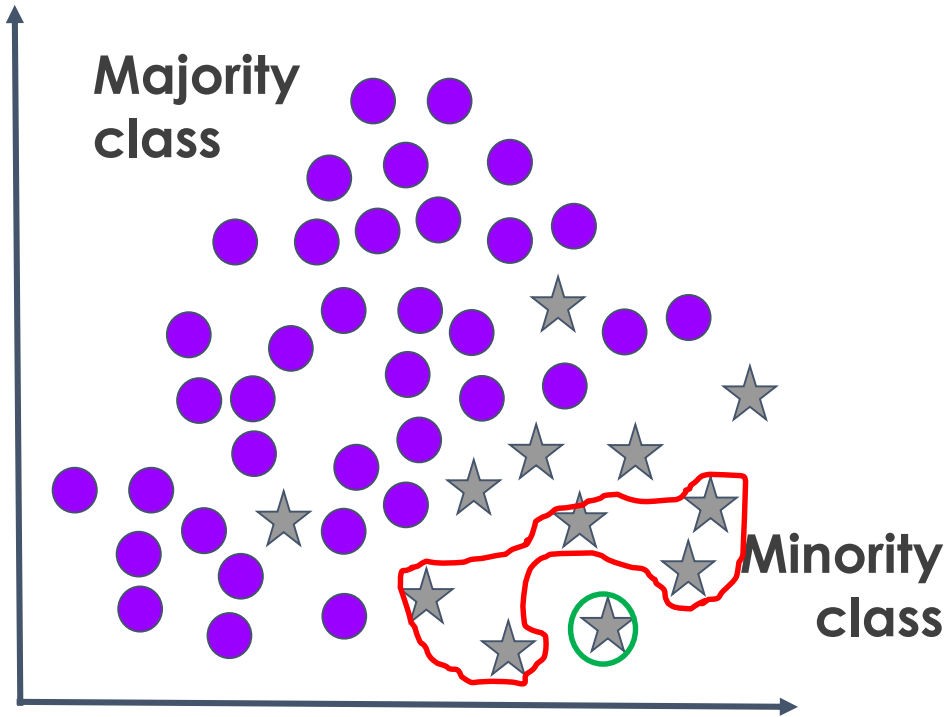
Borderline SMOTE: visually



Ignores noise:

Samples from minority class which neighbours are all from majority class.

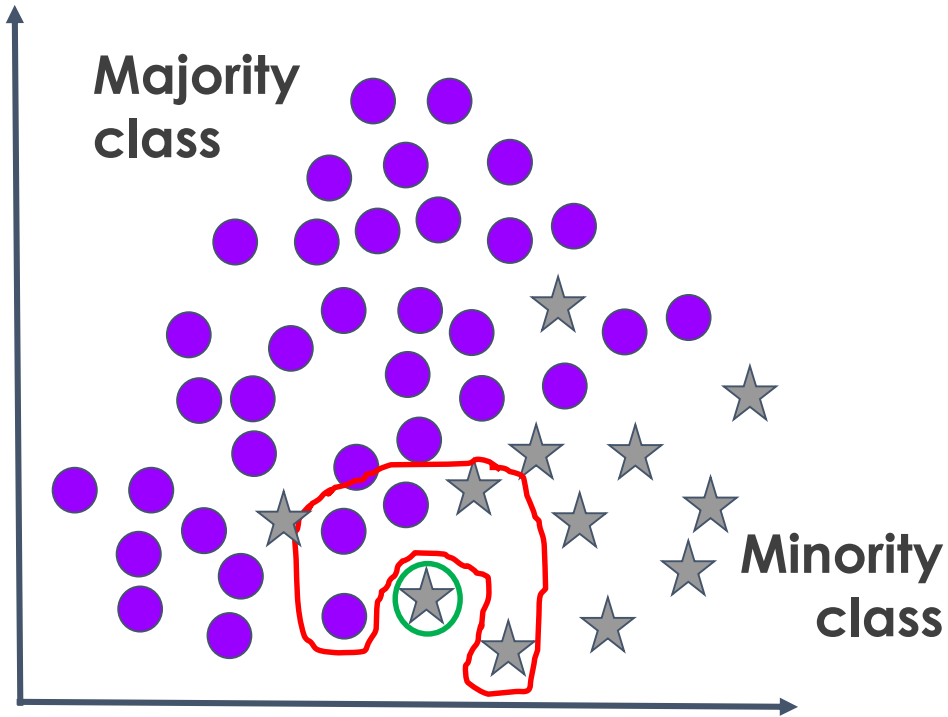
Borderline SMOTE: visually



Ignores safe group:

Samples from minority class which majority of neighbours are from minority class.

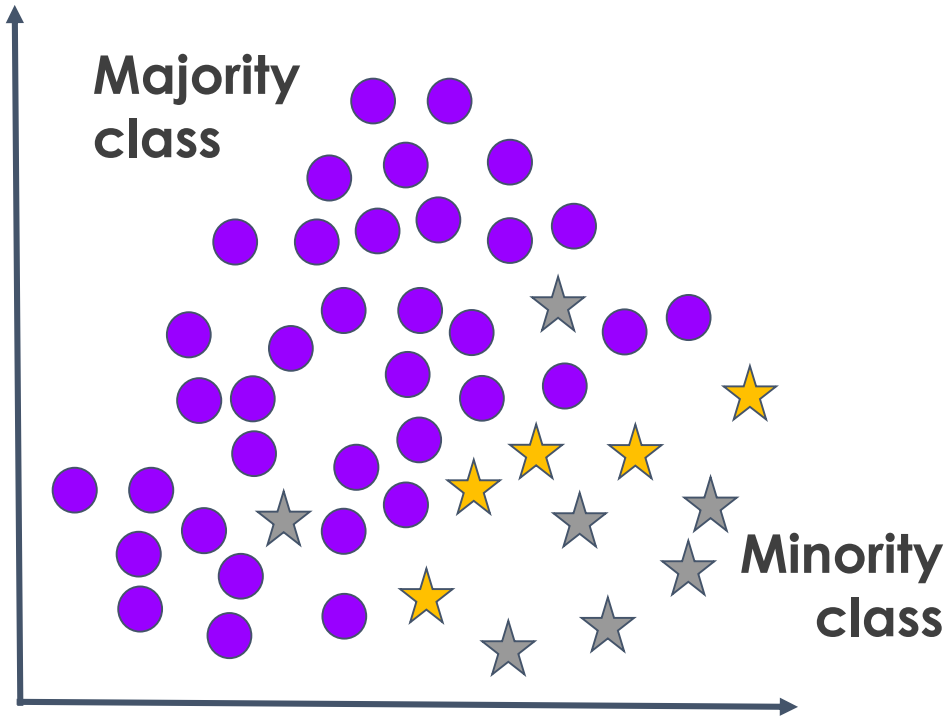
Borderline SMOTE: visually



Selects **DANGER** group:

Samples from minority class which majority of neighbours are from majority class.

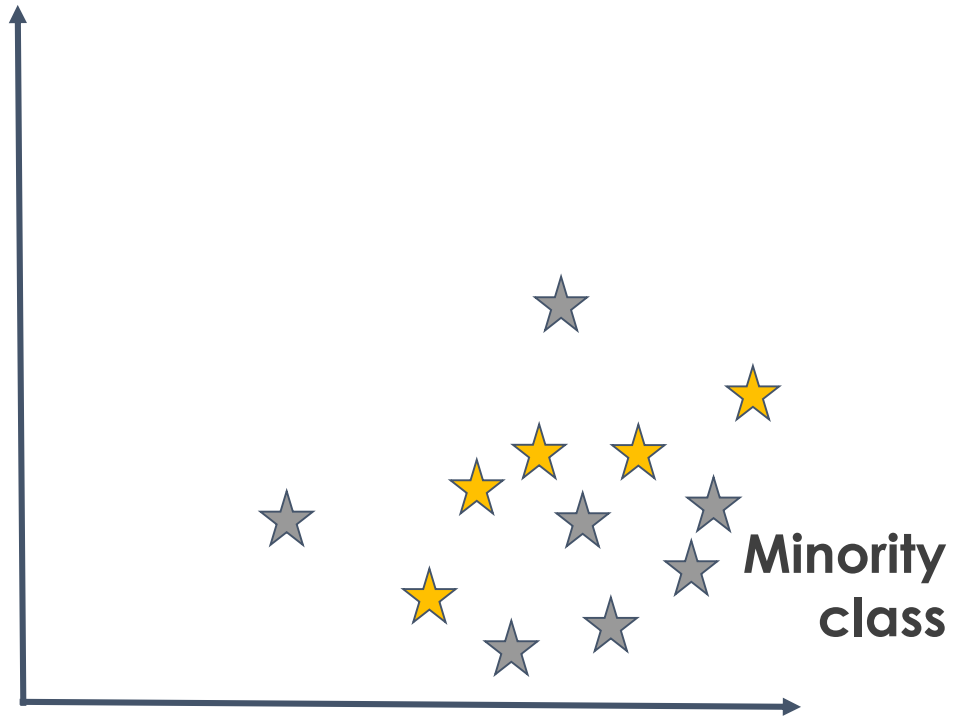
Borderline SMOTE: visually



Selects **DANGER** group:

Samples from minority class which majority of neighbours are from majority class.

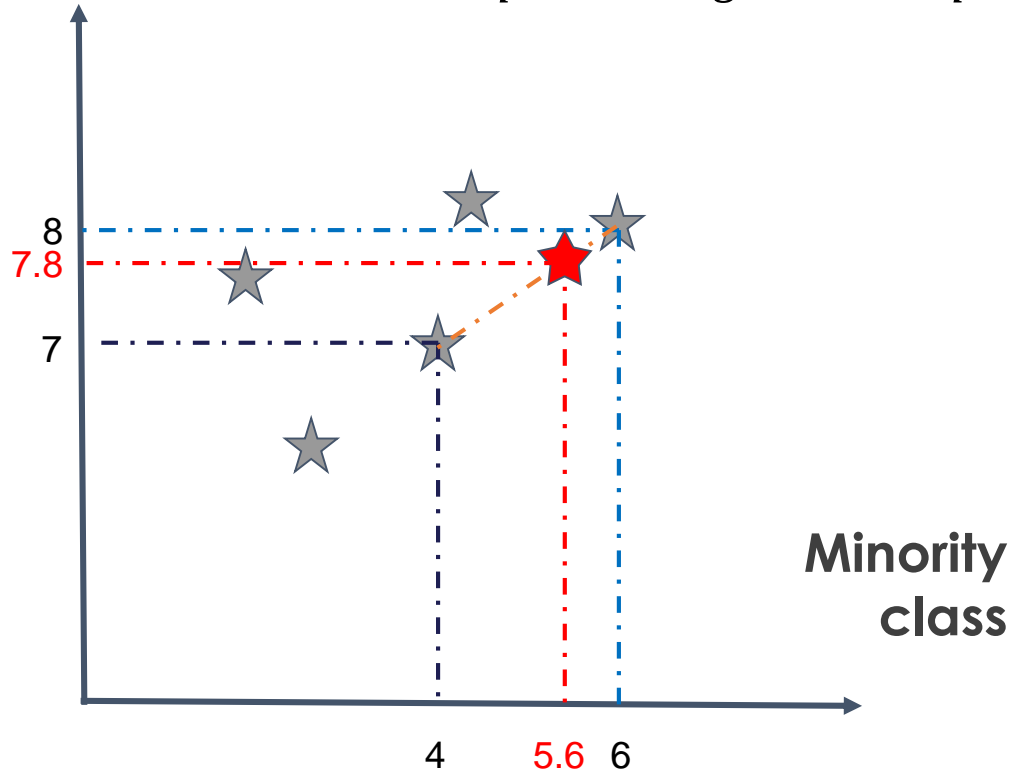
Borderline SMOTE: visual



Fits a KNN using minority class only (all examples)

Borderline SMOTE: variant 1

*New sample = original sample - factor * (original sample - neighbour)*



$$X_{\text{ori}} = (4, 7)$$

$$X_{\text{neig}} = (6, 8)$$

$$\text{New sample} = (4, 7) - 0.8 * ((4, 7) - (6, 8))$$

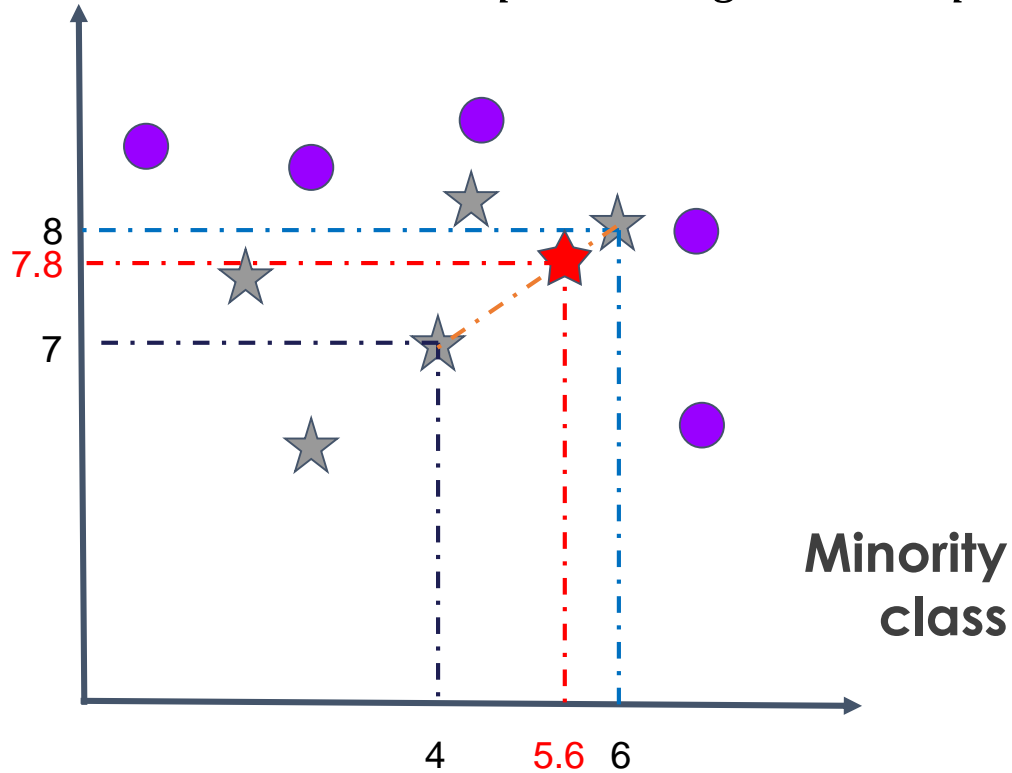
$$\text{New sample} = (4, 7) - 0.8 * ((-2, -1))$$

$$\text{New sample} = (4, 7) - ((-1.6, -0.8))$$

$$\text{New sample} = (5.6, 7.8)$$

Borderline SMOTE: variant 2

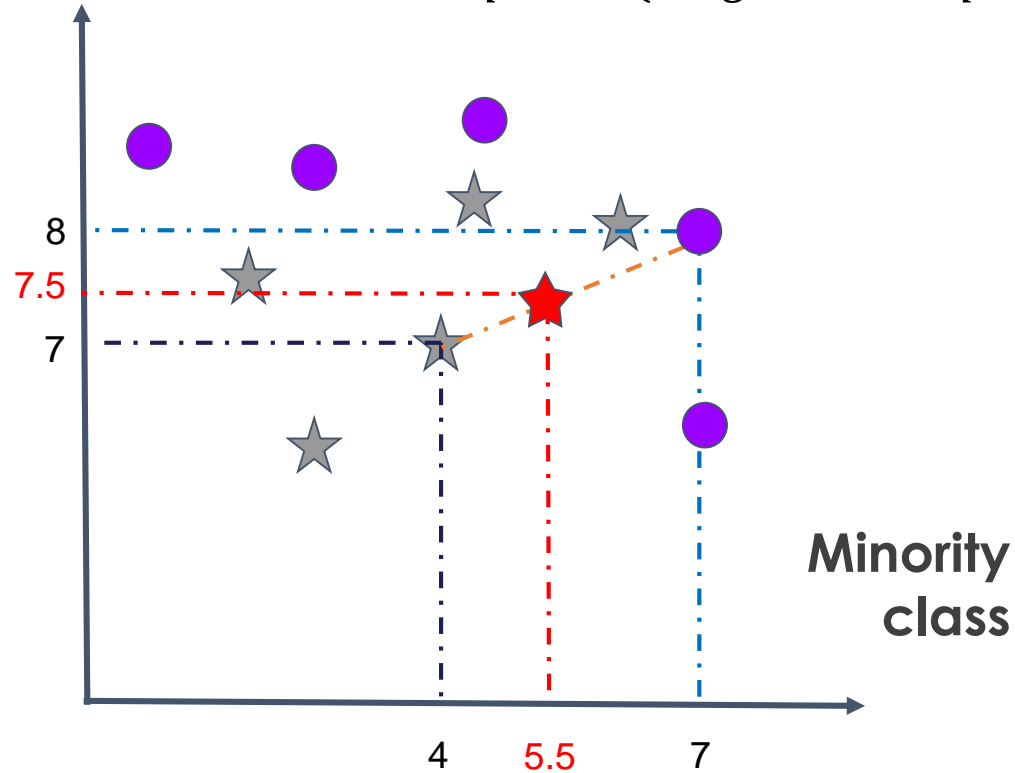
$$\text{New sample} = \text{original sample} - \text{factor} * (\text{original sample} - \text{neighbour})$$



Interpolation between DANGER observations and those of the same class is done as in variant 1

Borderline SMOTE: variant 2

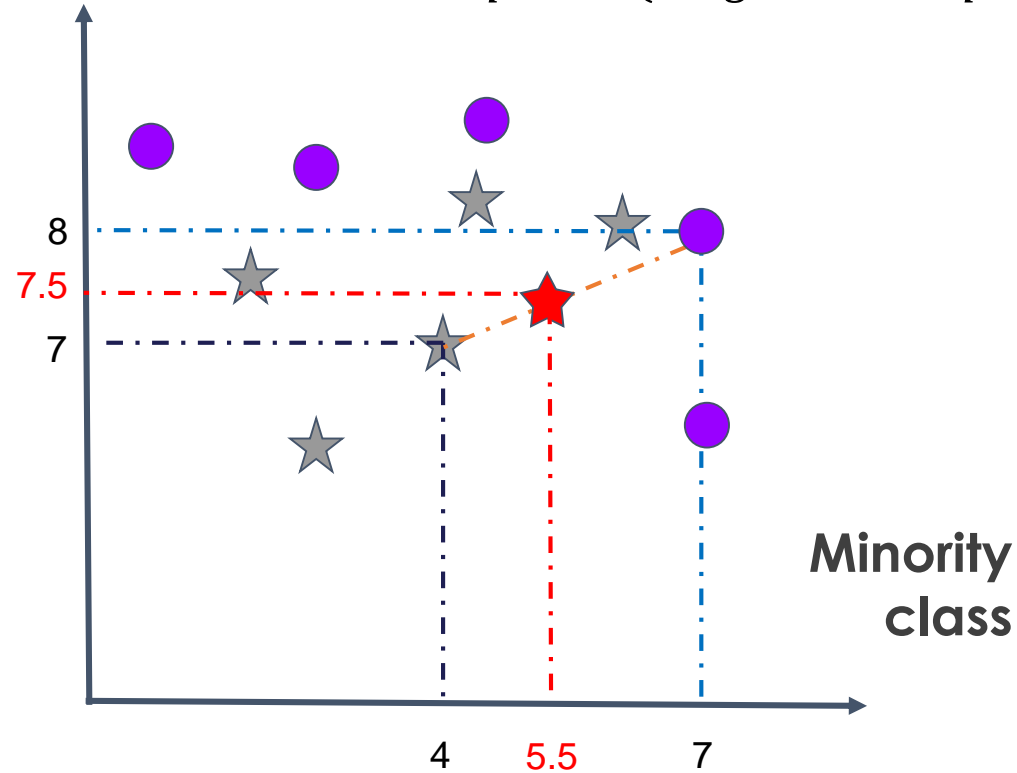
$$\text{New sample} = (\text{original sample} - \text{factor} * (\text{original sample} - \text{neighbour}))$$



Interpolation between
DANGER observations and
those of the majority class.

Borderline SMOTE: variant 2

$$\text{New sample} = (\text{original sample} - \text{factor} * (\text{original sample} - \text{neighbour}))$$



$$X_{\text{ori}} = (4, 7)$$

$$X_{\text{neig}} = (7, 8)$$

factor can take values between 0 and 0.5 instead of 0 and 1

$$\text{New sample} = ((4, 7) - 0.5 * ((4, 7) - (7, 8)))$$

$$\text{New sample} = ((4, 7) - 0.5 * ((-3, -1)))$$

$$\text{New sample} = ((4, 7) - ((-1.5, -0.5)))$$

$$\text{New sample} = (5.5, 7.5)$$

Imbalanced-learn: Borderline SMOTE

```
▶ sm_b1 = BorderlineSMOTE(  
    sampling_strategy='auto',  # samples only the minority class  
    random_state=0,  # for reproducibility  
    k_neighbors=5,  
    m_neighbors=10,  
    kind='borderline-1',  
    n_jobs=4  
)  
  
X_res_b1, y_res_b1 = sm_b1.fit_resample(X, y)
```


THANK YOU

www.trainindata.com