Train In Data

# Random Over-sampling with smoothing

# Random Over-sampling with smoothing

- Extracts observations at random from the minority class, and compounds its value with some noise.
- The noise is informed by the class distribution.
- Creates new examples ➔ Avoids data duplication.

# Random Over-sampling

| Target | Var 1 | Var 2 | Var 3 |
|---|---|---|---|
| 0 | 1 | 20 | 1.5 |
| 0 | 2 | 15 | 2 |
| 0 | 1.4 | 12 | 1.7 |
| 0 | 1.7 | 10 | 2 |
| 0 | 2 | 12 | 2 |
| 0 | 1.5 | 11 | 1.5 |
| 0 | 1 | 9 | 1 |
| 1 | 0.5 | 10 | 0.7 |
| 1 | 1 | 5 | 0.5 |
| 1 | 0 | 7 | 0.1 |

# Random Over-sampling

| Target | Var 1 | Var 2 | Var 3 |
|---|---|---|---|
| 0 | 1 | 20 | 1.5 |
| 0 | 2 | 15 | 2 |
| 0 | 1.4 | 12 | 1.7 |
| 0 | 1.7 | 10 | 2 |
| 0 | 2 | 12 | 2 |
| 0 | 1.5 | 11 | 1.5 |
| 0 | 1 | 9 | 1 |
| 1 | 0.5 | 10 | 0.7 |
| 1 | 1 | 5 | 0.5 |
| 1 | 0 | 7 | 0.1 |
| 1 | 0.5 | 10 | 0.7 |
| 1 | 1 | 5 | 0.5 |
| 1 | 0 | 7 | 0.1 |

Extract samples at random ➔ duplication

# Random Over-sampling with smoothing

| Target | Var 1 | Var 2 | Var 3 |
|---|---|---|---|
| 0 | 1 | 20 | 1.5 |
| 0 | 2 | 15 | 2 |
| 0 | 1.4 | 12 | 1.7 |
| 0 | 1.7 | 10 | 2 |
| 0 | 2 | 12 | 2 |
| 0 | 1.5 | 11 | 1.5 |
| 0 | 1 | 9 | 1 |
| 1 | 0.5 | 10 | 0.7 |
| 1 | 1 | 5 | 0.5 |
| 1 | 0 | 7 | 0.1 |
| 1 | 0.5 | 10 | 0.7 |
| 1 | 1 | 5 | 0.5 |
| 1 | 0 | 7 | 0.1 |

1. Extract samples at random
2. Compound their value with some “noise”
3. The noise contemplates the class distribution
4. We can choose how disperse we want the new samples ➔ shrinkage factor (arbitrary)

# Random Over-sampling with smoothing

| Target | Var 1 | Var 2 | Var 3 |
|---|---|---|---|
| 1 | 0.5 | 10 | 0.7 |
| 1 | 1 | 5 | 0.5 |
| 1 | 0 | 7 | 0.1 |

| | | |
|---|---|---|
| 0.5 | 2.52 | 0.31 |

Standard deviation of each variable

1. Take the minority samples and determine their distribution ➔ Standard deviation of each variable

# Random Over-sampling with smoothing

| Target | Var 1 | Var 2 | Var 3 |
|---|---|---|---|
| 1 | 0.5 | 10 | 0.7 |
| 1 | 1 | 5 | 0.5 |
| 1 | 0 | 7 | 0.1 |

| | | |
|---|---|---|
| 0.5 | 2.52 | 0.31 |

Standard deviation of each variable

| | | |
|---|---|---|
| 0.1 | -0.5 | 0.2 |

Random initial value

1. Take the minority samples and determine their distribution ➔ Standard deviation of each variable

2. Extract a value at random from N ~(0, 1)

# Random Over-sampling with smoothing

| Target | Var 1 | Var 2 | Var 3 |
|---|---|---|---|
| 1 | 0.5 | 10 | 0.7 |
| 1 | 1 | 5 | 0.5 |
| 1 | 0 | 7 | 0.1 |

| | | |
|---|---|---|
| 0.5 | 2.52 | 0.31 |

Standard deviation of each variable

| | | |
|---|---|---|
| 0.1 | -0.5 | 0.2 |

Random initial value

| | | |
|---|---|---|
| 1 | 5 | 0.5 |

1. Take the minority samples and determine their distribution ➔ Standard deviation of each variable
2. Extract a value at random from N ~(0, 1)
3. Extract at random 1 observation

# Random Over-sampling with smoothing

| Target | Var 1 | Var 2 | Var 3 |
|---|---|---|---|
| 1 | 0.5 | 10 | 0.7 |
| 1 | 1 | 5 | 0.5 |
| 1 | 0 | 7 | 0.1 |

| | | |
|---|---|---|
| 0.5 | 2.52 | 0.31 |

Standard deviation of each variable

| | | |
|---|---|---|
| 0.1 | -0.5 | 0.2 |

Random initial value

| | | |
|---|---|---|
| 1 | 5 | 0.5 |

1. Take the minority samples and determine their distribution ➔ Standard deviation of each variable
2. Extract a value at random from N ~(0, 1)
3. Extract at random 1 observation
4. Add the noise (with the shrinking factor, e.g. 10)

| | | |
|---|---|---|
| 1+(**10** * 0.5* 0.1) | 5+(**10** * 2.52* (-0.5)) | 0.5+(**10** * 0.31 *0.2) |

# Random Over-sampling with smoothing

| Target | Var 1 | Var 2 | Var 3 |
|---|---|---|---|
| 1 | 0.5 | 10 | 0.7 |
| 1 | 1 | 5 | 0.5 |
| 1 | 0 | 7 | 0.1 |

| | | |
|---|---|---|
| 0.5 | 2.52 | 0.31 |

Standard deviation of each variable

| | | |
|---|---|---|
| 0.1 | -0.5 | 0.2 |

Random initial value

| | | |
|---|---|---|
| 1 | 5 | 0.5 |

| | | | |
|---|---|---|---|
| New example | 1.5 | -7.6 | 1.2 |

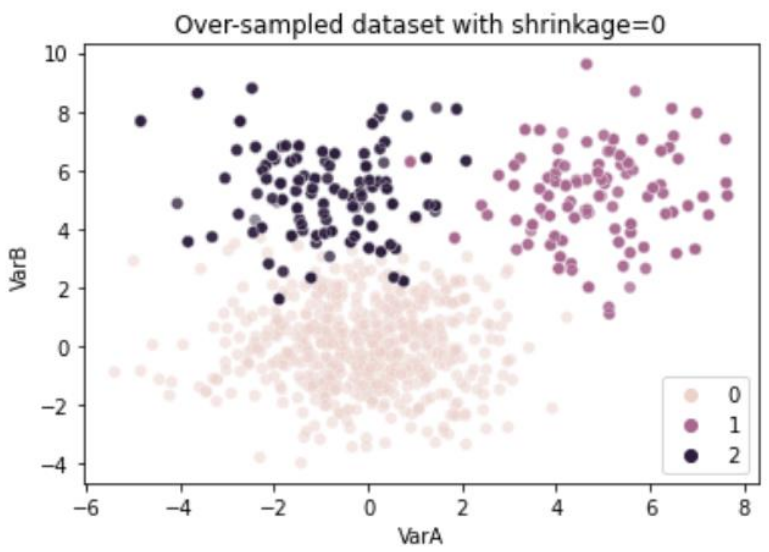
1. Take the minority samples and determine their distribution ➔ Standard deviation of each variable
2. Extract a value at random from N ~(0, 1)
3. Extract at random 1 observation
4. Add the noise (with the shrinking factor, e.g. 10)

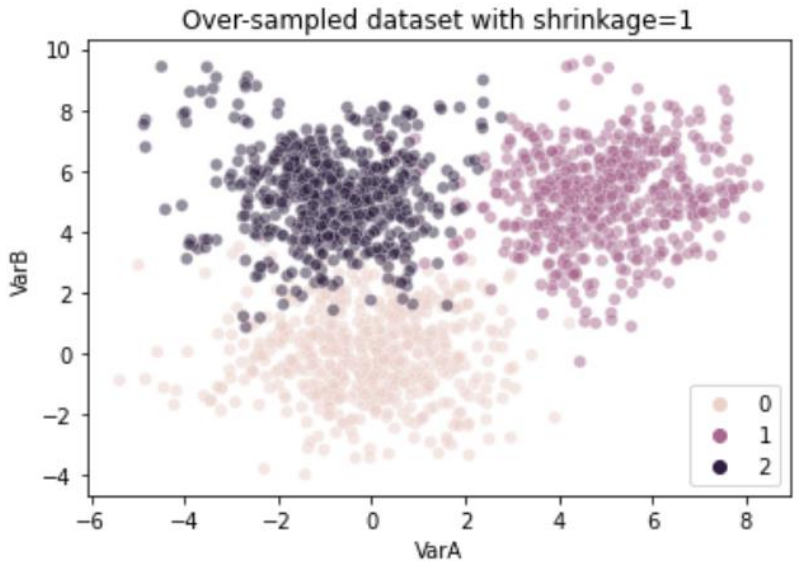
# Imbalanced-learn: RandomOverSampler

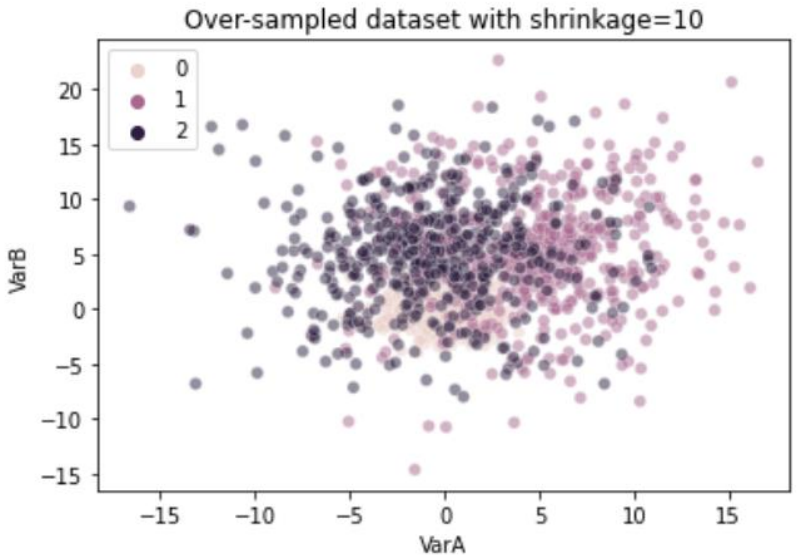
```
ros = RandomOverSampler(
       strategy = ‘auto’,
       random_state = 29,
       shrinkage = 10,
)

X_res, y_res = ros.fit_resample(X, y)
```

# Imbalanced-learn: RandomOverSampler




The bigger the shrinkage factor, the greater the dispersion of the final samples

Train In Data

# THANK YOU

www.trainindata.com