

## 1. Introduction

- **Propensity scores:** conditional probability of belonging in treatment group [1]
- Positivity assumption not enough
- **Density estimation:** estimate shape of distributions

## 2. Research Question

“When do propensity score methods with density estimation work well and when do they fail to identify overlap for different types of datasets?”

How do the following affect the performance?

- Classifiers
- Features
- Outliers

## 3. Method

- Estimating overlap region:

$$P(X = x|T = t) = \frac{P(T = t|X = x)P(X = x)}{P(T = t)}$$

- Classifiers: logistic regression, decision trees, random forests
- Kernel density estimation (KDE)
- Total overlap region kept around 30%

## 4. Results

A: 1 region of overlap

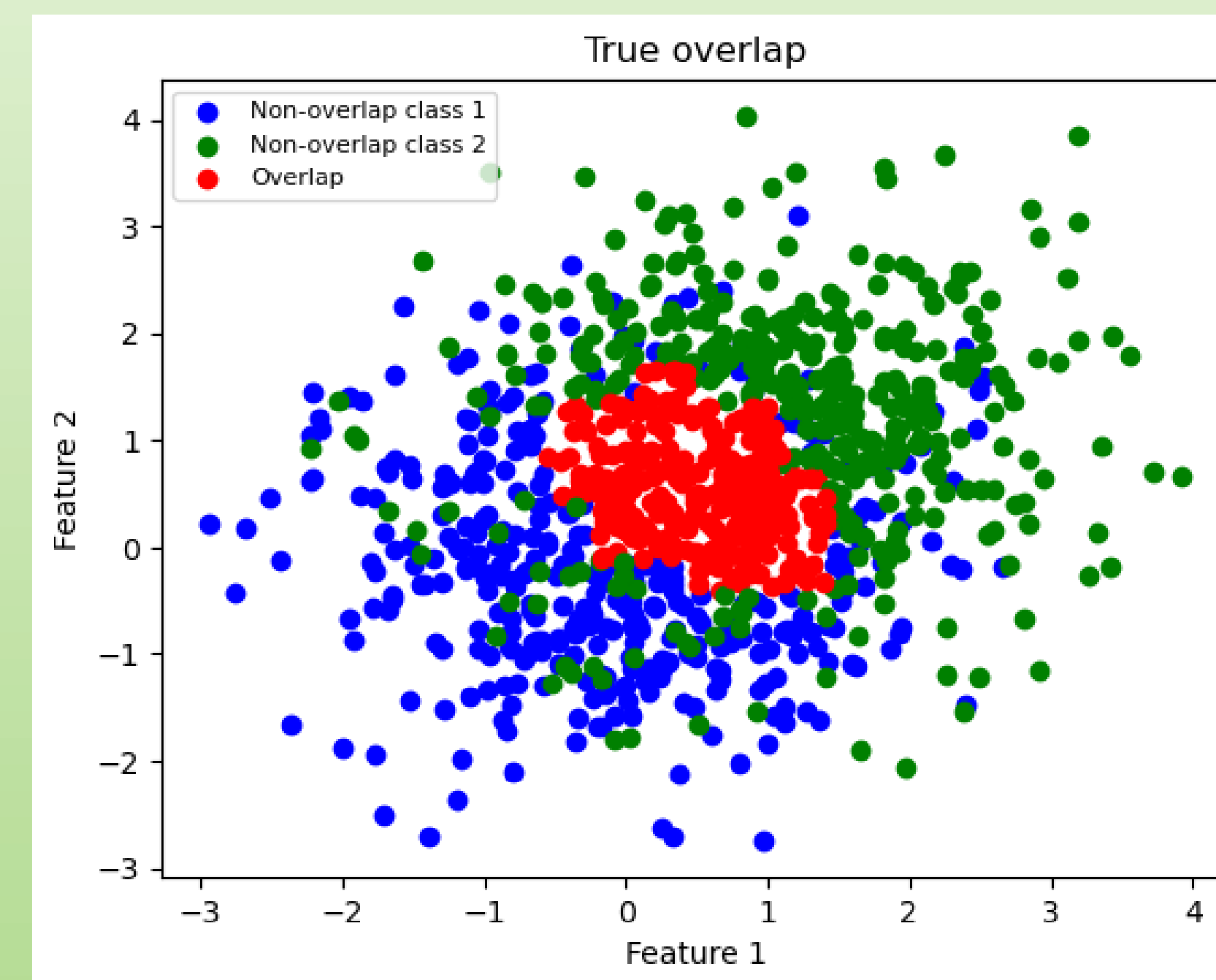


Figure 1: Example of true overlap for 1 region of overlap, with 2 features (2D).

B: 2 regions of overlap

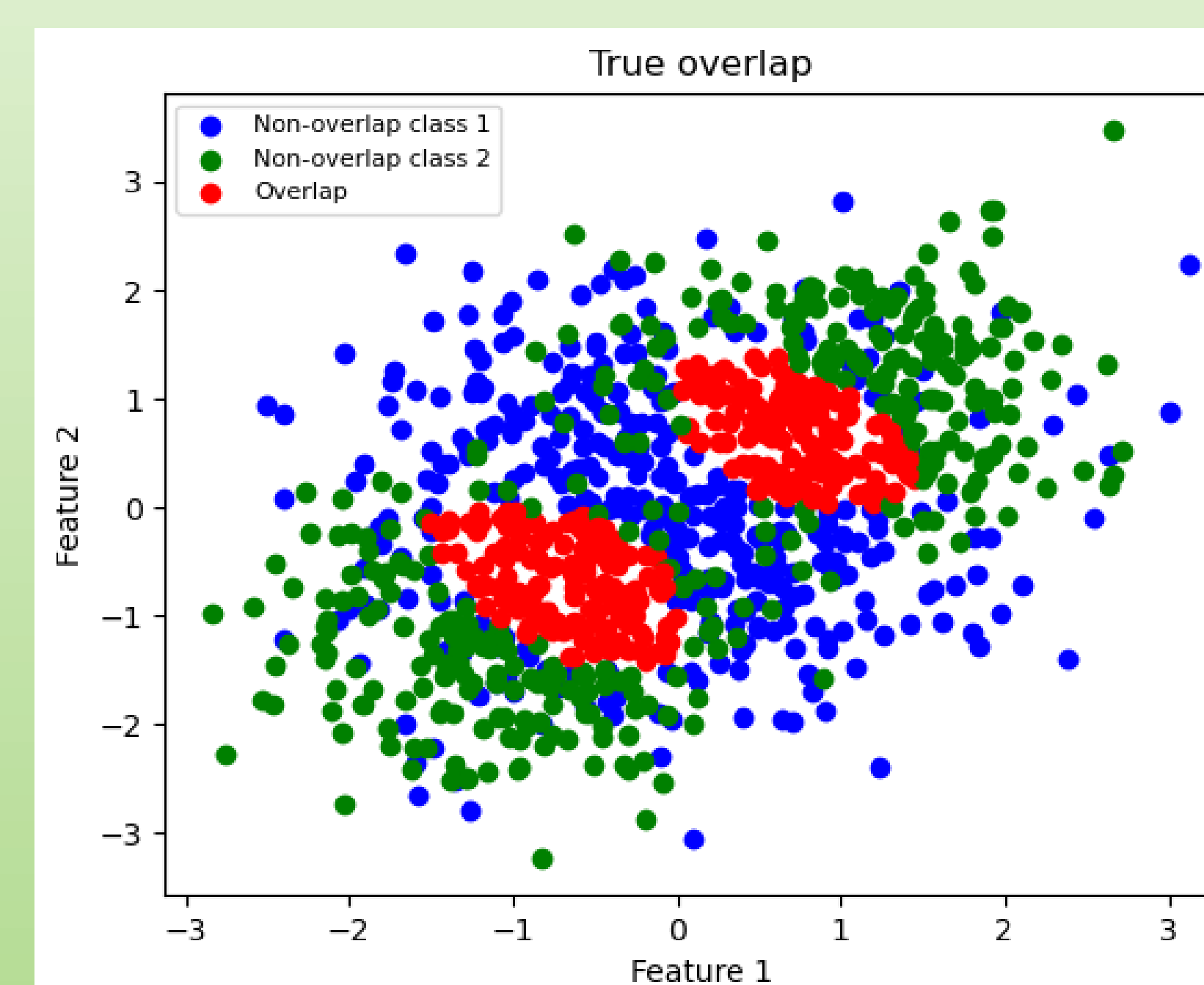


Figure 3: Example of true overlap for 2 regions of overlap, with 2 features (2D).

C: Outliers

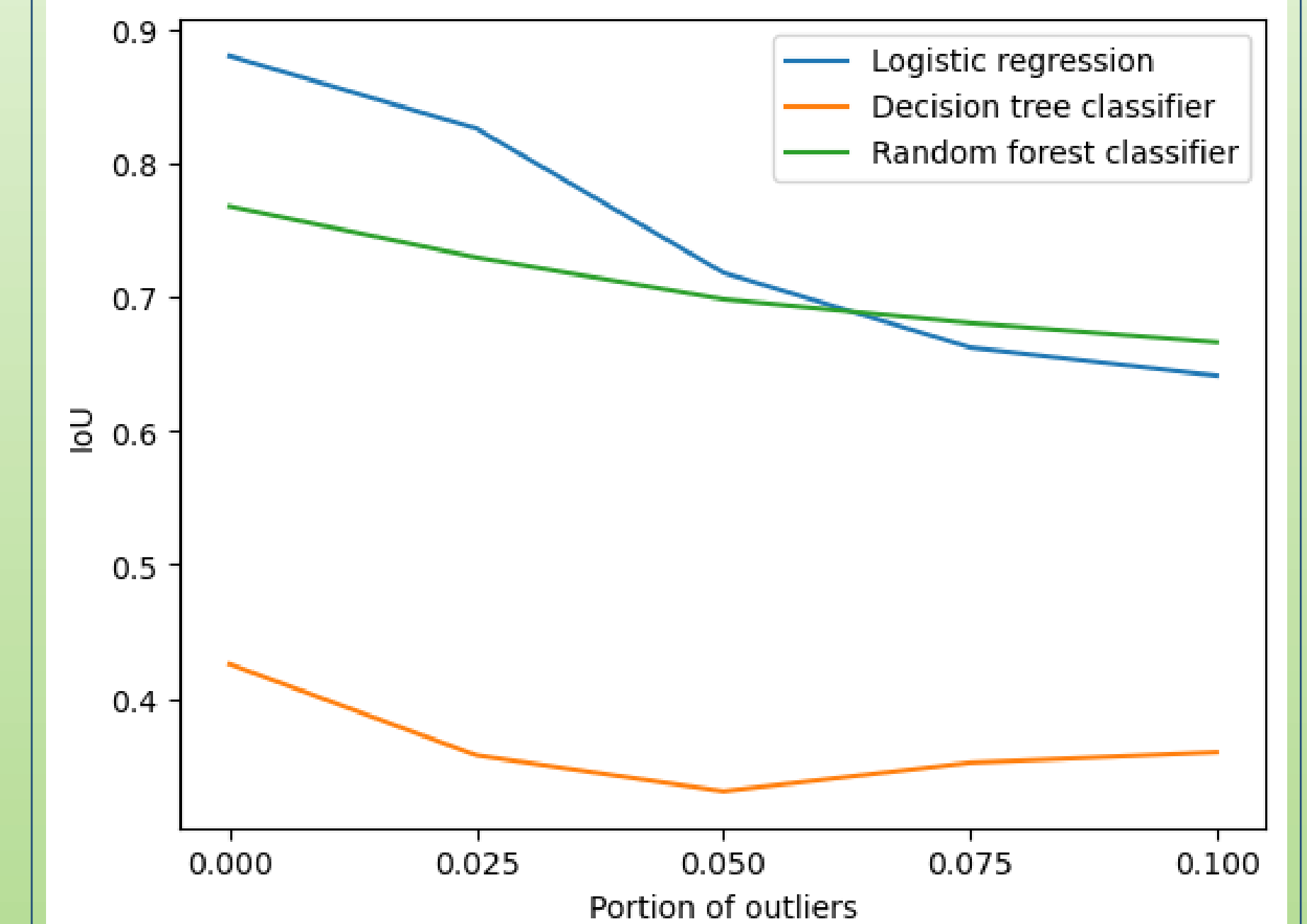


Figure 5: Performance with an increasing number of outliers. One region of overlap with 2 features (2D).

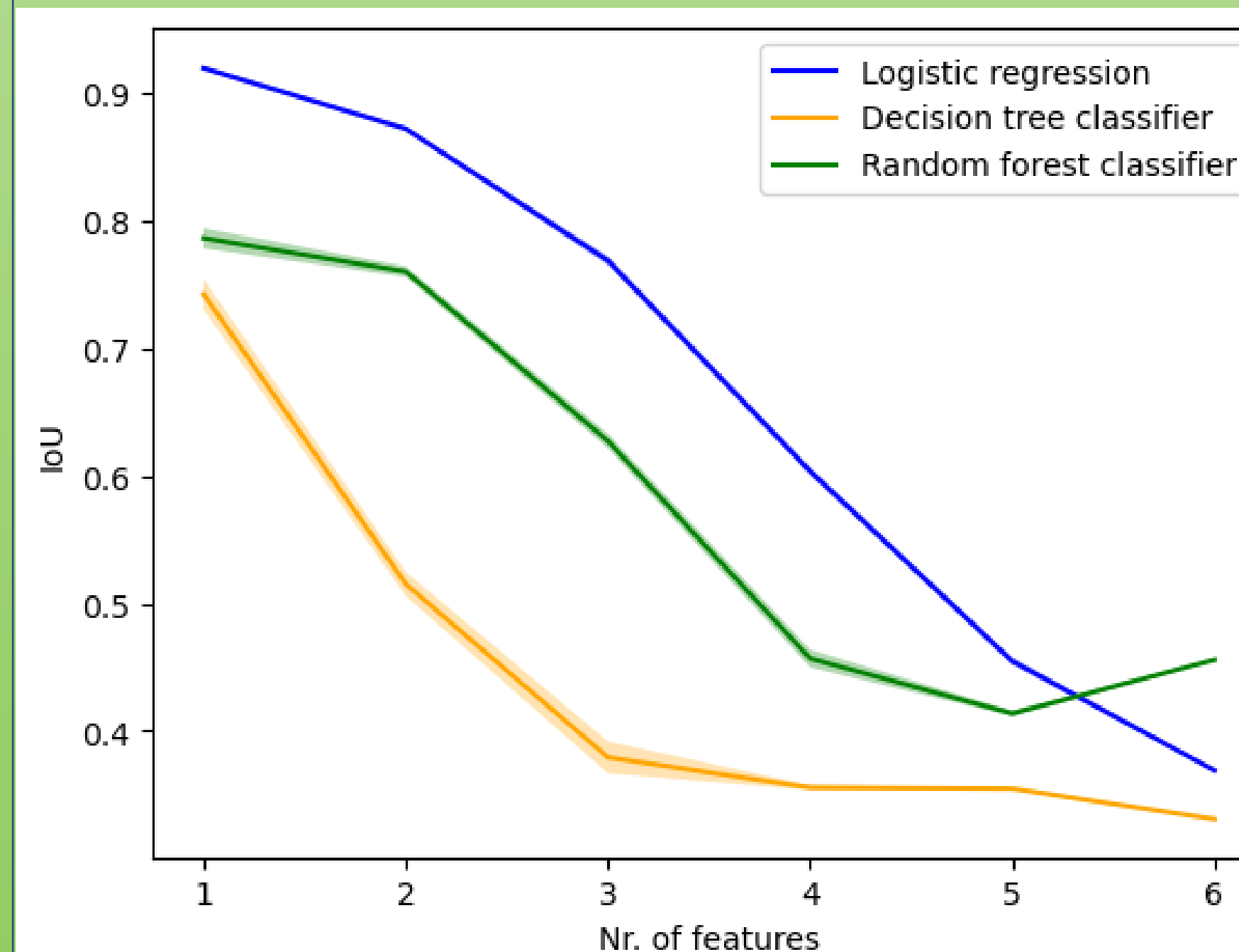


Figure 2: Performance for one region of overlap with an increasing number of features.

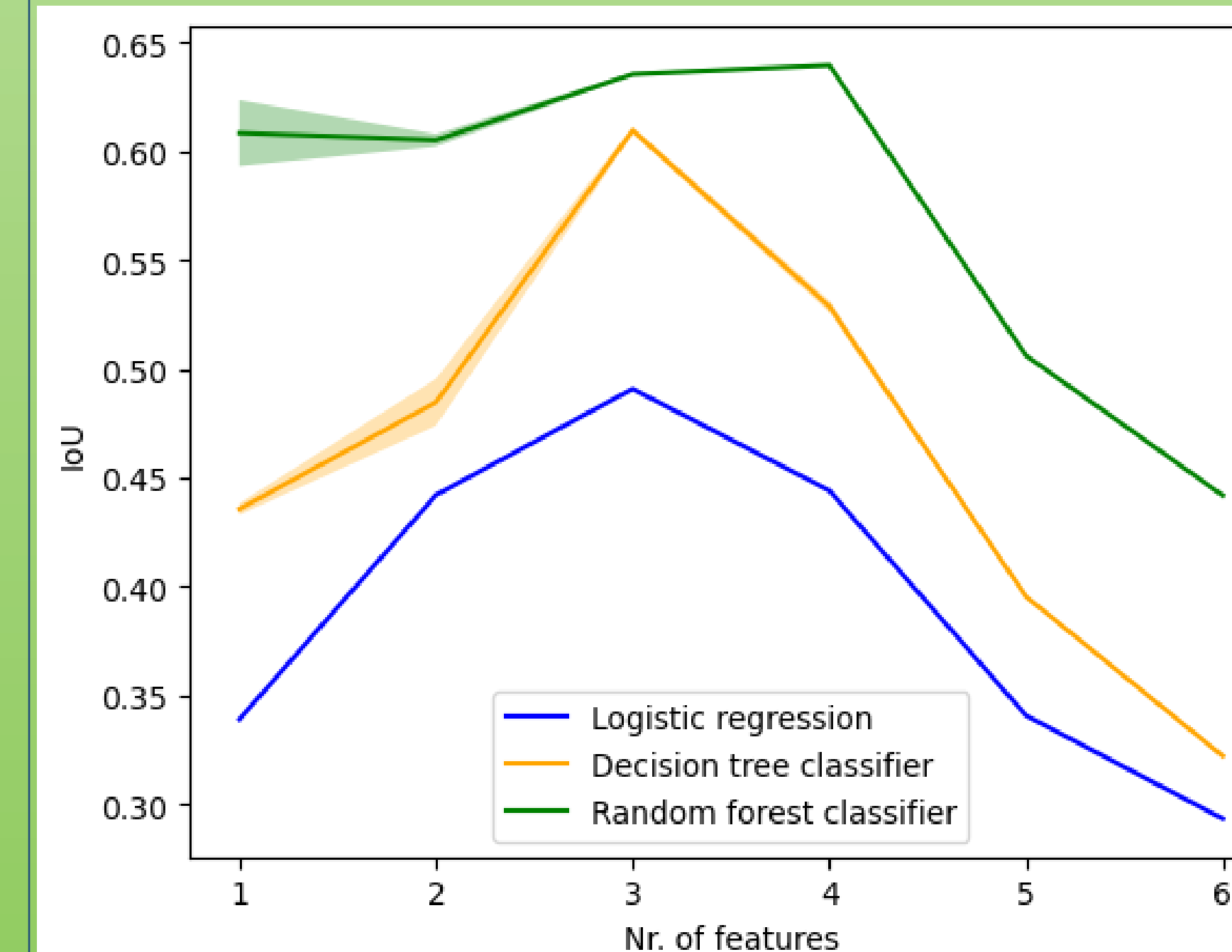


Figure 4: Performance for two regions of overlap with an increasing number of features.

## 5. Conclusion

- Performs well with 1 region of overlap, worse with multiple regions
- Higher dimensionality generally makes performance worse
- Slightly affected by up to 2.5% outliers. More outliers only affects logistic regression

## 6. Future work

- Observe behaviour with real datasets
- 3+ classes with multiclass classifiers