Anomaly Detection with Mixed Integer Optimization

Why do we care?

When we study data behavior and causal relationships, outliers often skew our models and estimations. Naive interpretation and generalization of models without outlier removal may result in misleading conclusions. Therefore, we consider identifying outliers a crucial preliminary action both for anomaly detection applications and for improving supervised learning accuracy.

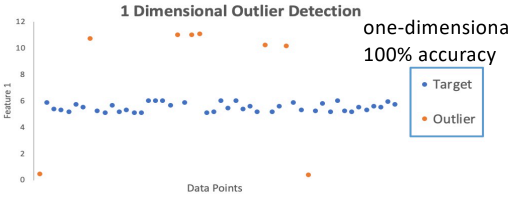
First Approach

We here propose an optimization perspective of solving the traditional machine learning problem:

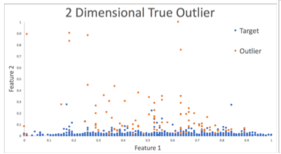
1. Use MIO to identify feature-wise outliers;
2. Study whether removing outliers form datasets can improve the modeling process. We hope the MIO approach on detecting outliers can be seen as a innovative approach to traditional problems.

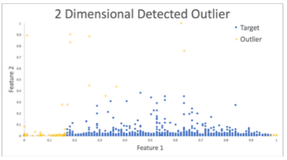
$$\begin{aligned} \min \quad & ||Xz - \text{median}(X)z||_2 \\ \text{s.t.} \quad & ||z||_0 \leq K \\ & z \in 0, 1 \end{aligned}$$

First Approach: Results



one-dimensional:
100% accuracy





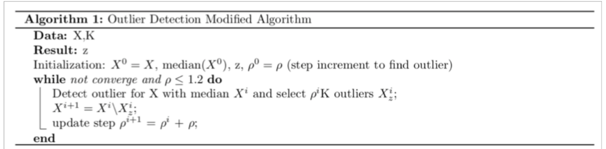
Failed to detect many of the outliers when the data is in a higher dimension

First Approach: Pros and Cons

- Pros: Intuitive, fast and interpretable, with high accuracy for unsupervised, and low-dimensional data.
- Cons: When data is high-dimensional or outliers are not evenly distributed with far Euclidean distance from median, this method does not capture such error.

First Approach: Modified

Simple outlier detection would not work for data with skewed median. We propose an alternate algorithm that keeps adjusting median while selecting outliers:



```
Algorithm 1: Outlier Detection Modified Algorithm
Data: X,K
Result: z
Initialization: $X^0 = X$, median($X^0$), z, $\rho^0 = \rho$ (step increment to find outlier)
while not converge and $\rho \leq 1.2$ do
    Detect outlier for X with median $X^i$ and select $\rho^i$K outliers $X^i_z$;
    $X^{i+1} = X^i \backslash X^i_z$;
    update step $\rho^{i+1} = \rho^i + \rho$;
end
```

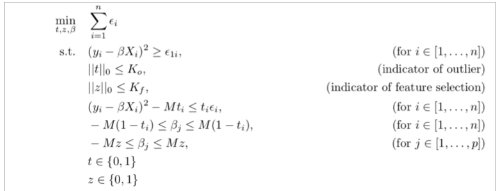
2 Dimensional True Outlier
Original Target
Detected Target
Original Outlier
Detected Outlier
Feature 2
Feature 1

Modified approach has Significantly improved model Accuracy.

Second Approach

Methodology:

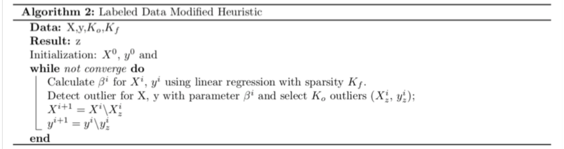
1. Detect outlier with labeled data, and minimize MSE for non-outlier observations.
2. Study outlier detecting and feature selection simultaneously:
3. Minimize mean squared error for selected observations with selected features.



$$\begin{aligned} \min_{t,z,\beta} \quad & \sum_{i=1}^{n} \epsilon_i \\ \text{s.t.} \quad & (y_i - \beta X_i)^2 \geq \epsilon_{1i}, && \text{(for } i \in [1, \ldots, n]) \\ & ||t||_0 \leq K_o, && \text{(indicator of outlier)} \\ & ||z||_0 \leq K_f, && \text{(indicator of feature selection)} \\ & (y_i - \beta X_i)^2 - M t_i \leq t_i \epsilon_i, && \text{(for } i \in [1, \ldots, n]) \\ & -M(1 - t_i) \leq \beta_j \leq M(1 - t_i), && \text{(for } i \in [1, \ldots, n]) \\ & -Mz \leq \beta_j \leq Mz, && \text{(for } j \in [1, \ldots, p]) \\ & t \in \{0, 1\} \\ & z \in \{0, 1\} \end{aligned}$$

Second Approach: Alternating Heuristic

However, with sample of 500 observations, it took more than 15 hours to solve in Julia. Thus we here propose heuristics to solve such problem:



```
Algorithm 2: Labeled Data Modified Heuristic
Data: X,y,$K_o$,$K_f$
Result: z
Initialization: $X^0$, $y^0$ and
while not converge do
    Calculate $\beta^i$ for $X^i$, $y^i$ using linear regression with sparsity $K_f$.
    Detect outlier for X, y with parameter $\beta^i$ and select $K_o$ outliers ($X^i_z$, $y^i_z$);
    $X^{i+1} = X^i \backslash X^i_z$
    $y^{i+1} = y^i \backslash y^i_z$
end
```

Second Heuristic Approach: Pros and Cons

- Pros: Improve outlier detection performance for high-dimensional data, and account for correlation within features.
- Cons: Computational expensive, and algorithm accuracy strongly correlates with model linear regression performance.

Final Results and Comparison

Data	Baseline Accuracy	Accuracy: First Approach	Accuracy: First Approach Modified
throid	97.5%	95.8%	99.9%
cardio	90.3%	92.1%	91.8%
iris	66.7%	93.3%	100%

Table 1: Anomaly Detection

Data	MSE: Original	MSE: First Approach	MSE: Second Heuristic	Improve: First Approach	Improve: Second Heuristic
housing	30	25.4	22.3	15.40%	25.7%
crime	0.016	0.013	0.018	18.70%	-11.1%
cancer	573.8	571.8	519.4	0.30%	9.5%
Insurance	4.18*10e7	4.18*10e7	4.2*10e7	0%	-0.47%
housing 2	13664	13663	13661	0.01%	0.02%

Table 2: Application in predictive models

Conclusion

1. Outlier removal improves model performance in most cases, but not always.
2. Simultaneously removing outliers and fitting linear regression improve model performance further.
3. Real-life anomaly observations may not distant significantly from targeting data points.
4. MIO approach focuses on overall outlier removal, whereas in real-life, outliers may appear in only 1 feature.
5. Feature correlation can impede outlier detection.

Challenges and Future Improvements

1. It is hard to decide on the number of outliers that we want to remove from the datasets
2. When data lies in a higher dimension, the accuracy decreases.
3. Measuring distance from median to detect outliers can be ineffective.
4. For future improvement, instead of only using distance to detect outliers, more criterion can be taken into consideration.

Vicky Liu