We explored the effect of 3 classes of linear transformations: Trend extraction (e.g., candlestick charting, EMA), Feature selection (e.g., low variance, random forest selection), and Dimensionality reduction (e.g., PCA).

- For instance, [3] presents theoretical evidence that datasets with higher intrinsic dimensionality facilitate adversarial attacks.
- Adjacently, the presentation of data to a learning model impacts its performance. For example, we have seen this through the use of dimensionality reduction over the years to increase model accuracy.
- Adversarial research has focused primarily on classification problems in computer vision applications.

### Background

- Current adversarial machine learning research continuously approaches similar conclusions: the vulnerability of ML models is highly correlated to how the data is represented [1, 2].
  - For instance, [3] presents theoretical evidence that datasets with higher intrinsic dimensionality facilitate adversarial attacks.
- To avoid introducing additional vulnerabilities in ML pipelines, ML practitioners must observe and understand the particular dataset’s intrinsic characteristics and ensure any transformation does not stray away from the intrinsic dimension.

### Problem

ML practitioners in time series fields may unknowingly making more vulnerable models with the use of certain data transformations.

### Research Aims

We designed our experimentation to address the following:
1. Could data transformations contribute to any adversary's ability to more easily construct adversarial examples?
2. Is the dimensionality reduction technique, PCA, consistent as a strategy to increase robustness when given a time series dataset, RNN, and varying selected principal components?
3. What representations of time series data contribute to ML models that are least susceptible to adversarial examples?

We explored the effect of 3 classes of linear transformations:
- Dimensionality reduction (e.g., PCA)
- Feature selection (e.g., low variance, random forest selection)
- Trend extraction (e.g., candlestick charting, EMA)

### Evaluation Results

<table>
<thead>
<tr>
<th>Data Transformation</th>
<th>Feature Count</th>
<th>Baseline Accuracy</th>
<th>Distance (δ)</th>
<th>Δ in Robustness</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA 50%</td>
<td>11</td>
<td>97.90%</td>
<td>0.51</td>
<td>24.39%</td>
</tr>
<tr>
<td>PCA 81%</td>
<td>18</td>
<td>98.80%</td>
<td>0.76</td>
<td>44.90%</td>
</tr>
<tr>
<td>PCA 27%</td>
<td>6</td>
<td>95.00%</td>
<td>0.34</td>
<td>66.98%</td>
</tr>
<tr>
<td>Random Forest</td>
<td>9</td>
<td>96.11%</td>
<td>0.13</td>
<td>31.71%</td>
</tr>
<tr>
<td>Low Variance</td>
<td>11</td>
<td>91.32%</td>
<td>0.15</td>
<td>65.85%</td>
</tr>
<tr>
<td>Candlesticks</td>
<td>22</td>
<td>92.78%</td>
<td>0.11</td>
<td>60.98%</td>
</tr>
<tr>
<td>EMA</td>
<td>22</td>
<td>96.48%</td>
<td>0.51</td>
<td>7.32%</td>
</tr>
</tbody>
</table>

The most robust RNN used PCA with only half of the principal components and is only a consistent defense against adversarial examples if the number of selected principle components approximates the data's intrinsic dimension.

### Discussion

- PCA creates more well-defined sub-manifolds for each class such that an adversary requires higher perturbation to "trick" an ML model.
  - This is not the case for feature selection and trend extraction since there is no mapping to a lower dimension.

### References


