CDMC2020 Task Presentation

Rongxin Xu, Xiaoyan Wang

2020-10-11

Overview

Cross-validation
visualization
One-Class Classification

Cross-validation

- PCA-30D
- T-SNE

visualization

- T-SNE
- PCA

One-Class Classification

- One-Class SVM
- Variational Auto-Encoder

Cross-validation
PCA-30D
T-SNE
visualization
One-Class Classification

Cross-validation

PCA-30D

Defn Use PCA to reduce the data to 30 dimensions, and then Use 5 fold cross-validation to check accuracy scores.

Table 1: PCA-30D accuracy score

	fold1	fold2	fold3	fold4	fold5	mean
XGBoost	96.167%	95.998%	96.078%	96.238%	96.389%	96.174%
LightGBM	96.026%	95.740%	95.887%	95.933%	96.118%	95.941%
Randomfoest	96.392%	96.183%	96.299%	96.416%	96.564%	96.371%

T-SNE

Cross-validation
PCA-30D
T-SNE
visualization
One-Class Classification

Defn Predict the category of malware based on a given byte sequence.

Table 2: T-SNE accuracy score

	fold1	fold2	fold3	fold4	fold5	mean
XGBoost	91.226%	91.609%	91.267%	91.669%	91.378%	91.430%
LightGBM	88.636%	88.611%	88.639%	88.636%	88.541%	88.613%
Randomfoest	94.465%	94.810%	94.611%	94.776%	94.538%	94.640%

Cross-validation
visualization
T-SNE
PCA
One-Class Classification

visualization

T-SNE

Cross-validation
visualization
T-SNE
PCA
One-Class Classification

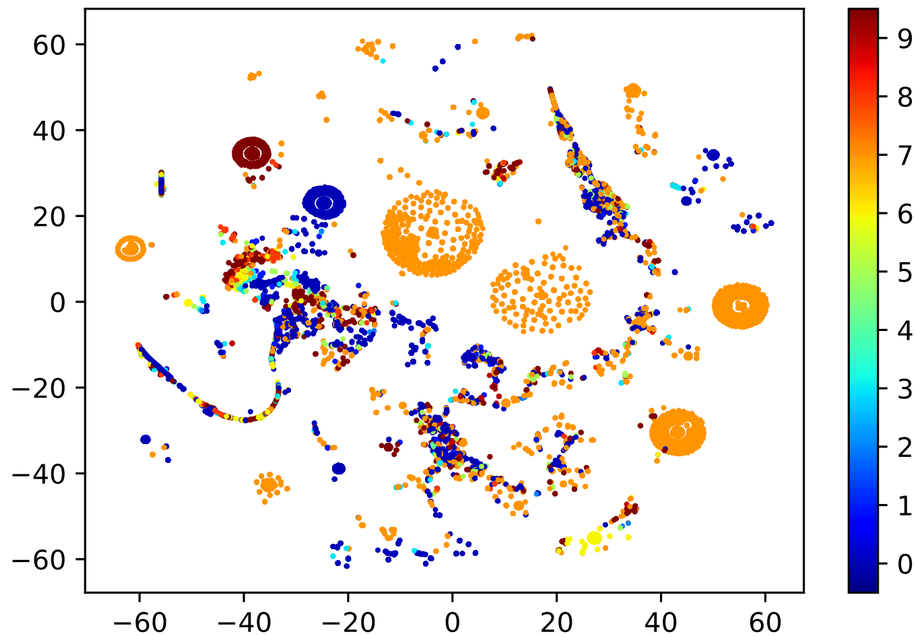

Figure 1: T-SNE

PCA

Cross-validation
visualization
T-SNE
PCA
One-Class Classification

The PCA method using maximum likelihood estimation can only be reduced to 798 dimensions. The image below is a visualization image reduced to 2 dimensions using PCA.

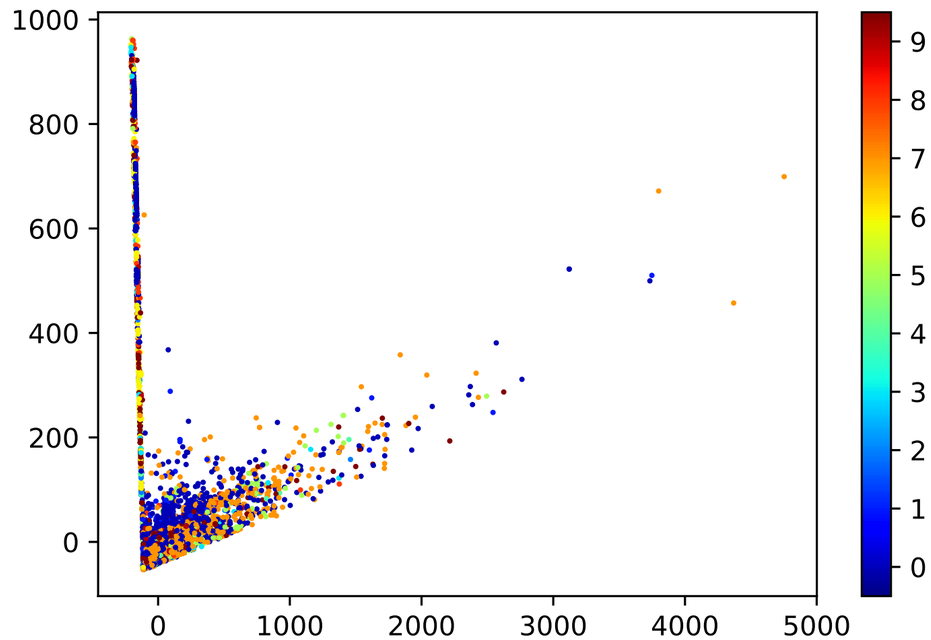

Figure 2: pca-2D

Cross-validation
visualization
One-Class Classification
One-Class SVM
Variational Auto-Encoder

One-Class Classification

One-Class SVM

Cross-validation
visualization
One-Class Classification
One-Class SVM
Variational Auto-Encoder

	0	1	2	3	4	5	6	7	8	9	10	11
0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
2	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
3	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
4	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
...	...	...	...	...	...	...	...	...	...	...	...	...
6275	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
6276	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
6277	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
6278	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
6279	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1

Variational Auto-Encoder

Cross-validation
visualization
One-Class Classification
One-Class SVM
Variational Auto-Encoder

- Hierarchial VAE
- Other models?

End

Thank you!