Class 1

Class 2

[Yang et al.: "A Closer Look at Accuracy vs. Robustness", 2020]
\[ \varepsilon_{\text{min}} = r \]

[Yang et al.: „A Closer Look at Accuracy vs. Robustness”, 2020]
MSCR… Minimal Separation Corruptions Robustness

MSCR = \frac{Acc_{robust} - Acc_{original}}{Acc_{original}}

100% Acc_{original} > Acc_{robust} \Rightarrow MSCR < 0

100% Acc_{original} = 100% Acc_{robust} \Rightarrow MSCR = 0
MSCR Metric Applicability

Robustness over Training Noise on CIFAR-10

Trained for higher robustness

Class Separation Distance for Robustness Evaluation of Classifiers
Accuracy-Robustness-Tradeoff?

2D Dataset

Optimum

Robustness [MSCR]

Clean Accuracy

97% 98% 99% 100%

-0.15 -0.10 -0.05 0.00 0.05 0.10

0.03 0.02 0.015 0.002 0.001

Model with $\epsilon_{\text{train}} = \epsilon_{\text{min}}$

Baseline Model
MSCR metric:
• High interpretability

Accuracy-Robustness-Tradeoff:
• Not inherent in our experiments
• Sweetspots via Data Augmentation
Get into contact:

siedel.georg@baua.bund.de

Backup
2D data: 1NN vs Random Forest

- Robustness [MSCR] for 1NN vs. RF over training noise on 2D Dataset
  - MSCR RF
  - MSCR 1NN

- Clean Accuracy [%] for 1NN vs. RF over training noise on 2D Dataset
  - Clean Accuracy RF
  - Clean Accuracy 1NN

Class Separation Distance for Robustness Evaluation of Classifiers

Date: 25.07.2022
Accuracy-Robustness-Tradeoff?

CIFAR-10 Dataset

Robustness [MSCR]

Clean Accuracy

80% 85% 90% 95%

Model with $\varepsilon_{\text{train}} = 0.01$

Baseline Model

Optimum
CIFAR-10: Optima of Training vs. Test Noise