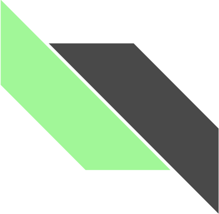

Attacks

18

- Current thinking on robust algorithms is to develop specific defensive mechanisms

- There are many variants of attack:

- Black-box attacks uses sampling to optimize a solution that fools the classifier

- White-box is where underlying structure is accessed

Attacks

- Current thinking on robust algorithms is to develop specific defensive mechanisms
- There are many variants of attack:
    - Black-box attacks uses sampling to optimize a solution that fools the classifier
    - White-box is where underlying structure is accessed

Uncertainty is the best form of defence

https://github.com/MadryLab/mnist_challenge, Sadeghi et al. 2019