Adversarial Training (Regular versus On-Manifold)

Model creation

Adversarial Training

Event Log

Train prefixes
- Prefix 1
- Prefix 2
- Prefix 3
- Prefix 4

Test prefixes
- Prefix 5
- Prefix 6
- Prefix 7

Classifier

Predictions*
(own training data)
- Prefix 3
- Prefix 1
- Prefix 2
- Prefix 4

adversarial attacks

adversarial attacks

Train prefixes and adversarial prefixes
- Prefix 3
- Prefix 2a
- Prefix 4a
- Prefix 4a
- Prefix 1
- Prefix 2
- Prefix 3
- Prefix 4

Classifier (adversarial)

Train prefixes (attacked)
- Prefix 3b
- Prefix 1b
- Prefix 2b
- Prefix 4b

Project on manifold

Train prefixes and on-manifold adversarial prefixes
- Prefix 1
- Prefix 2
- Prefix 3
- Prefix 4
- Prefix 3b
- Prefix 1b
- Prefix 2b
- Prefix 4b

Classifier (on-manifold)