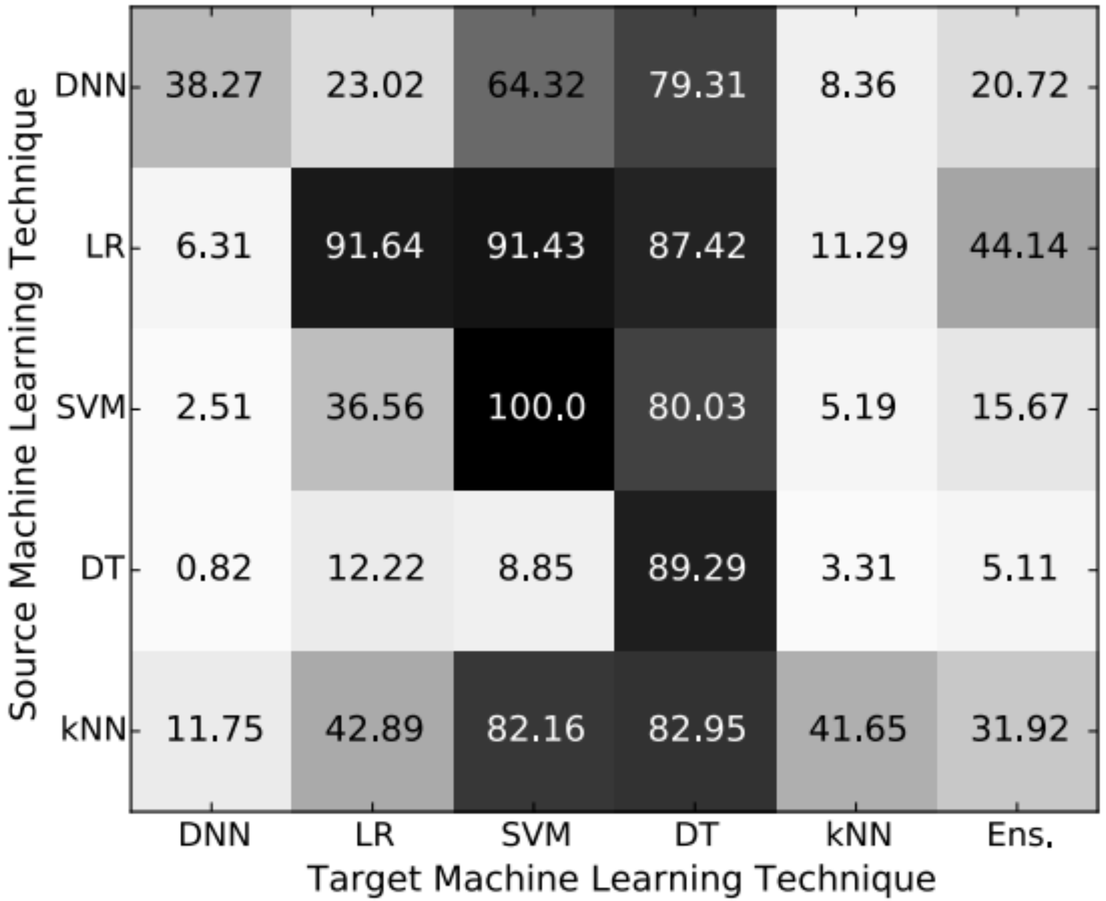
ADVERSARIAL MACHINE LEARNING

ATTACK MODELS : TRANSFERABILITY

- Transferability : Adversarial Examples crafted using one type of ML models can be used to attack other types of models
- This phenomena is observed because of similarity in decision boundaries of various ML models
- In The Space of Transferable Adversarial Examples F. Tramèr et al. analyse adversarial subspace.
- This is very important in devising Black Box Attacks

25

ATTACK MODELS: TRANSFERABILITY



Papernot, N., McDaniel, P., & Goodfellow, I. (2016). Transferability in machine learning: from phenomena to black-box attacks using adversarial samples.

ATTACK MODELS : TRANSFERABILITY

- Transferability : Adversarial Examples crafted using one type of ML models can be used to attack other types of models
- This phenomena is observed because of similarity in decision boundaries of various ML models
- In The Space of Transferable Adversarial Examples F. Tramèr et al. analyse adversarial subspace.
- This is very important in devising Black Box Attacks