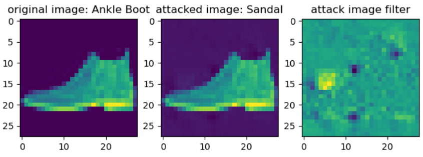
Attacks and Defenses

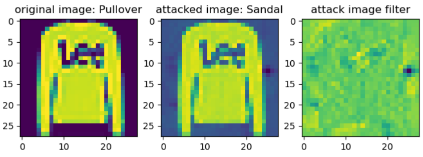
Caspar Schneider, Sinan Kuscu, Sinan Harms

September 2022

PART II - PRACTICE

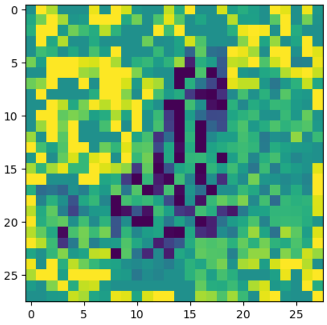
Task 1





- for fast attacks, classes that are more similar need less iterations (left: two iterations)
- the higher the difference between classes the more iterations are needed (right: ten iterations)
- similar classes have similar feature representations (see UMAP Sheet 2) therefore fast attacks are easier to perform

Task 2


0
5
10
15
20
25
0
5
10
15
20
25