

Top 10



Bottom 10

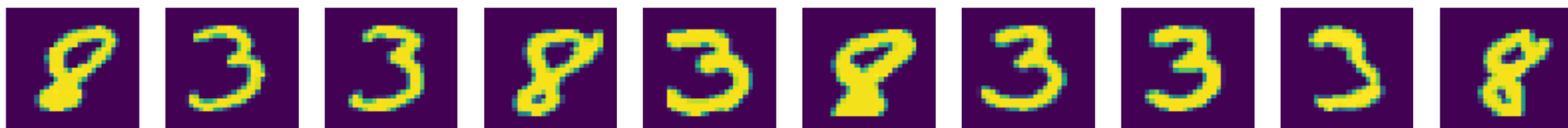


Figure 3. MNIST training digits sorted by norm of the removal update $\|\mathbf{H}_{\mathbf{w}^}^{-1}\Delta\|_2$. The samples with the highest norm (**top**) appear to be atypical, making it harder to undo their effect on the model. The samples with the lowest norm (**bottom**) are prototypical 3s and 8s, and hence are much easier to remove.*