Semi-supervised Learning (SSL)

- In SSL, one is given labeled and unlabeled data.
- Goal: train a classifier.

Motivation for Input and Weight Space Smoothing

- For many tasks like image classification, very small adversarial perturbations can make neural networks vulnerable to such perturbations.
- Over-parameterized networks are more robust to adversarial noises in the weight space even when they have the same decision boundary (i.e., the same input smoothness).

Input Smoothing and Weight Smoothing do not Imply Each Other.

- Input Smoothing: Approximates the loss with a 2nd order Taylor approximation of [1], and Approximate the Hessian with 1 step of power iteration.
- Weight Smoothing: Approximates the loss with the current estimate do not Imply Each Other.

Input Smoothing

Supervised Setting

\[
\Delta x = \arg \max \sum_{i \in X} \frac{f_i(x, w)}{f_i(x + \Delta x, w)}
\]

Joint Input and Weight Smoothing

\[
\Delta x = \arg \max \sum_{i \in X} \frac{f_i(x, w)}{f_i(x + \Delta x, w)}
\]

A New Algorithm for Weight Smoothing: Adversarial Block Coordinate Descent (ABCD)

Comparison to State-of-the-art.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CIFAR10</td>
<td>10.55</td>
<td>11.29</td>
<td>12.16</td>
<td>12.31</td>
<td>9.22</td>
<td>9.28</td>
<td>9.21 ± 0.21</td>
</tr>
<tr>
<td>SVHN</td>
<td>3.88</td>
<td>NR</td>
<td>4.42</td>
<td>5.88</td>
<td>3.95</td>
<td>3.55</td>
<td>3.53 ± 0.24</td>
</tr>
</tbody>
</table>


Robustness to Perturbations in Weight Space

Hessian of the Solution Converged