要点

- disagreement
- 更新不在只基于label，而是基于disagreement+label

方法

Algorithm 1 Update by Disagreement

input:
an update rule U
batch size b
two initial predictors $h_1, h_2 \in \mathcal{H}$
for $t = 1, 2, \ldots, N$ **do**
draw mini-batch $(x_1, y_1), \ldots, (x_b, y_b) \sim \tilde{\mathcal{D}}^b$
let $S = \{(x_i, y_i) : h_1(x_i) \neq h_2(x_i)\}$
$h_1 \leftarrow U(h_1, S)$
$h_2 \leftarrow U(h_2, S)$
end for

直观解释：

在训练开始阶段，两个网络的disagreement area 较大，就可以轻松的fit easy pattern而不会拟合噪声(memorization effect)

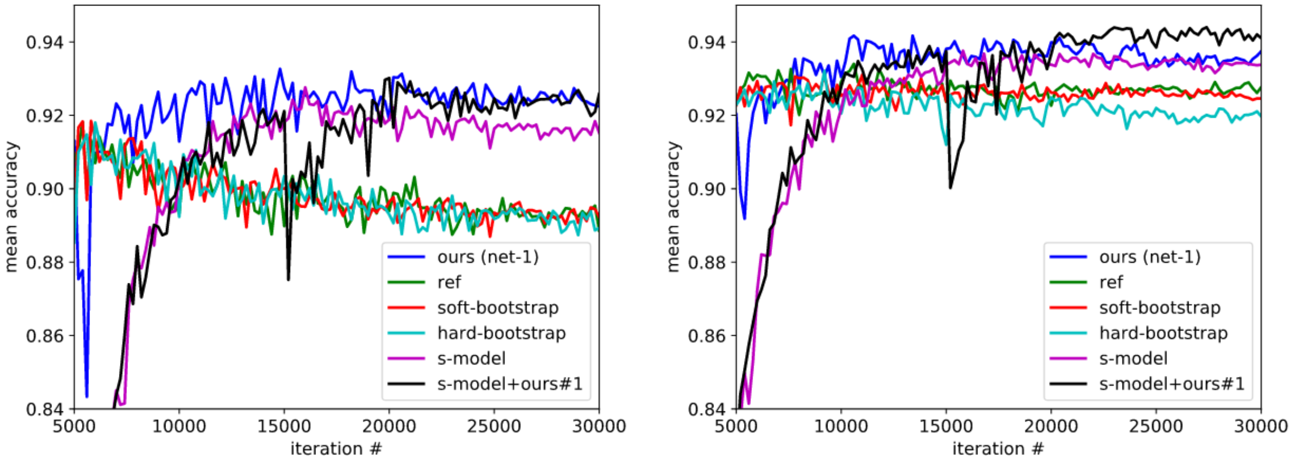
拟合完成easy pattern 之后， 理想情况下disagreement area 为 0 （因为两个网络都预测一个instance为一个标签，即使他是wrong label也不会被更新）

文章还给出了：

*disagreement 这种方法会收敛吗？ （收敛的很快）会收敛到最优吗？ （无法证明）*这两个问题的证明

实验

t%e



Dataset #1 - more noise

Dataset #2 - less noise

Figure 2: Balanced accuracy of all methods on clean test data, trained on the two different datasets.

先分别正常训练两个网络各5000轮，再执行decoupling。

（数据集是LFW）

Disagreement" rule. Due to the fact that we are not updating on all examples, we decrease the weight of batches with less than 10% of the original examples in the original batch to stabilize gradients. [2].

没看懂什么意思