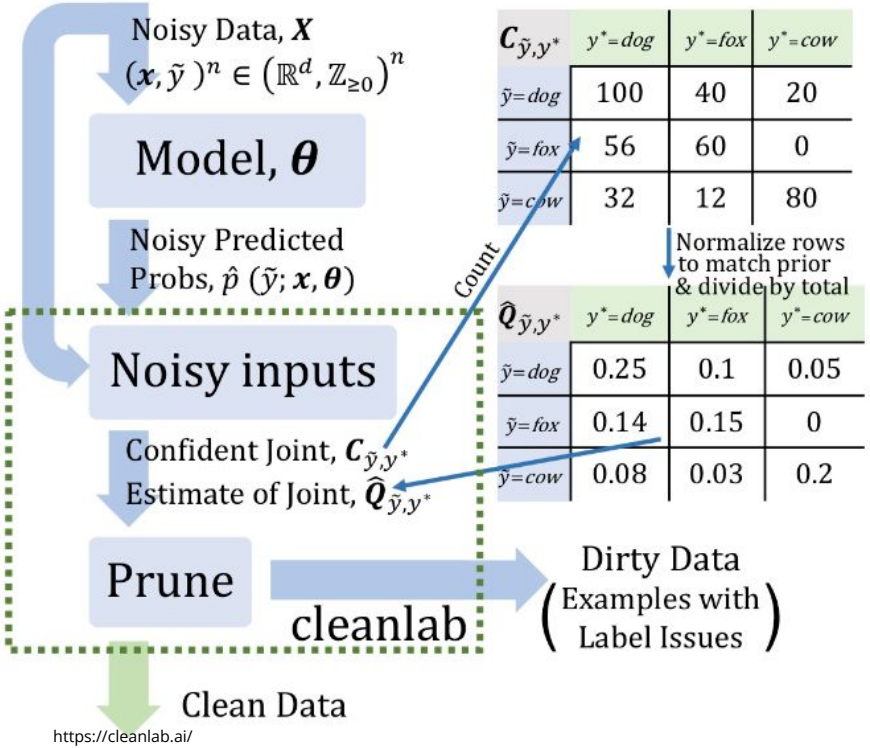
Confident Learning applied to MNIST Label Error

David Szczecina

Confident Learning and Cleanlab



$\boldsymbol{C}_{\tilde{y},y^*}$	$y^*=dog$	$y^*=fox$	$y^*=cow$
$\tilde{y}=dog$	100	40	20
$\tilde{y}=fox$	56	60	0
$\tilde{y}=cow$	32	12	80

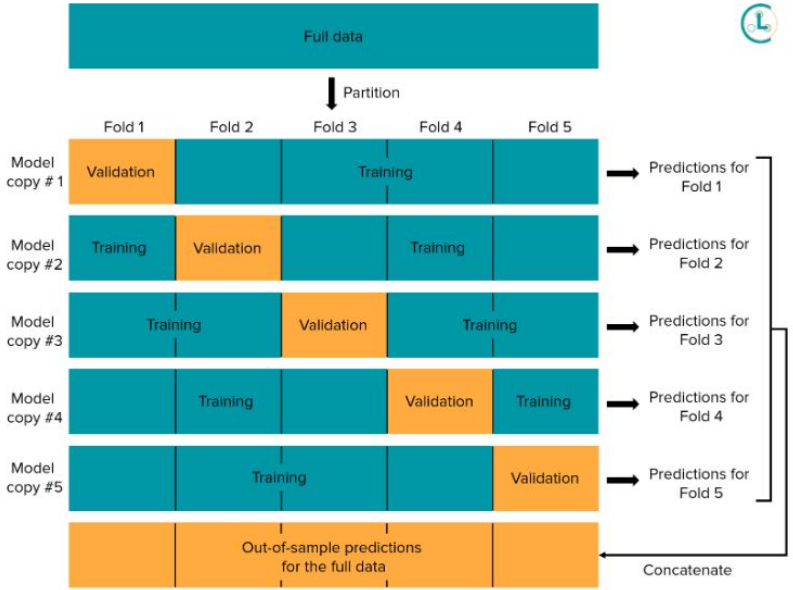
$\widehat{\boldsymbol{Q}}_{\tilde{y},y^*}$	$y^*=dog$	$y^*=fox$	$y^*=cow$
$\tilde{y}=dog$	0.25	0.1	0.05
$\tilde{y}=fox$	0.14	0.15	0
$\tilde{y}=cow$	0.08	0.03	0.2

1. Estimate the joint distribution of given, noisy labels and latent (unknown) uncorrupted labels to fully characterize class-conditional label noise.
2. Find and prune noisy examples with label issues.
3. Train with errors removed, re-weighting examples by the estimated latent prior.

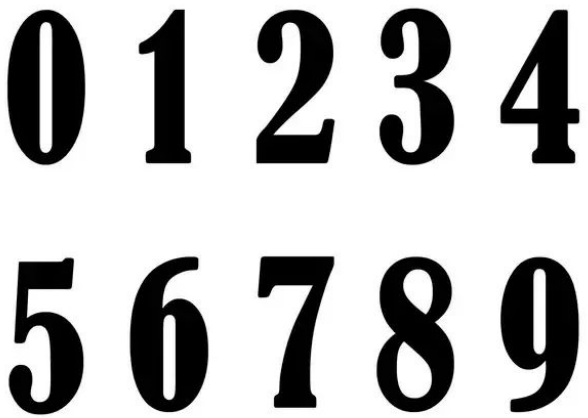
https://arxiv.org/pdf/1911.00068.pdf

Confident Learning and Cleanlab

Out of sample Predictions + Labels (can contain label errors)

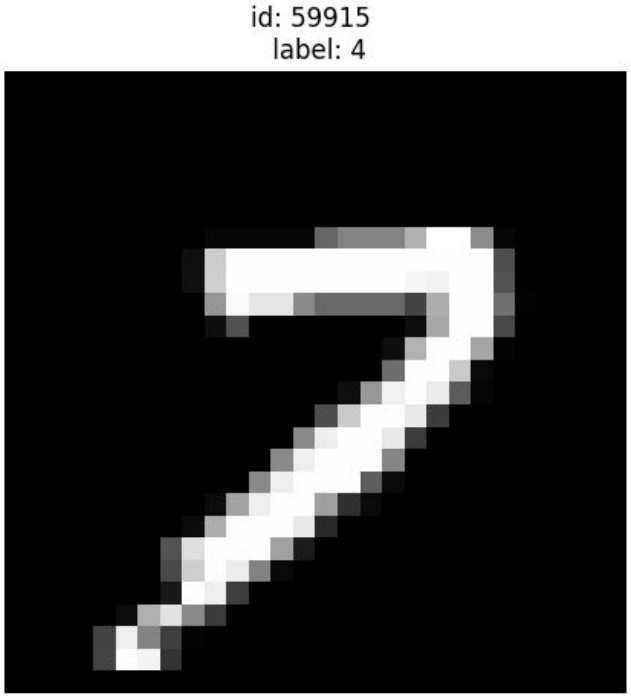


https://cleanlab.ai/



MNSIT Example

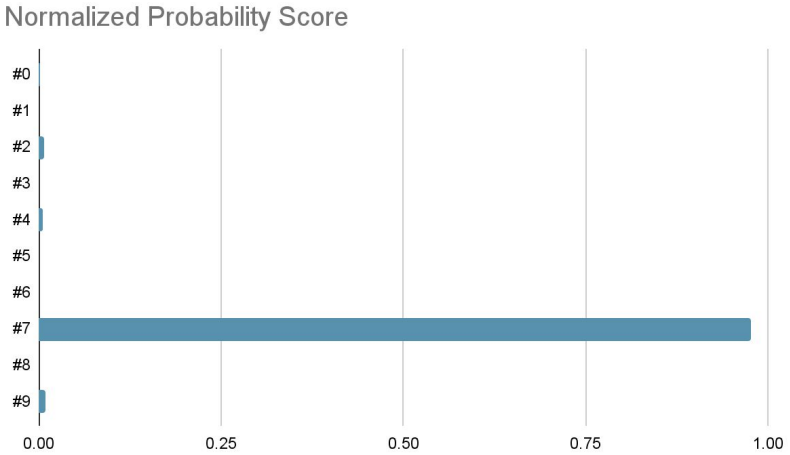
```
plot_examples([59915])
```



$$C_{\tilde{y},y^*}[i][j] := |\hat{X}_{\tilde{y}=i,y^*=j}| \quad \text{where}$$

$$\hat{X}_{\tilde{y}=i,y^*=j} := \left\{ x \in X_{\tilde{y}=i} : \hat{p}(\tilde{y}=j;x,\theta) \geq t_j,\ j = \underset{l \in [m]:\hat{p}(\tilde{y}=l;x,\theta) \geq t_l}{\arg\max} \hat{p}(\tilde{y}=l;x,\theta) \right\}$$

$$t_j = \frac{1}{|X_{\tilde{y}=j}|} \sum_{x \in X_{\tilde{y}=j}} \hat{p}(\tilde{y}=j;x,\theta)$$



*AVG scores [1, 0, 5, 0, 4, 0, 0, 737, 0, 7],

Noise Distribution

Actual

Labeled

```
[6837,    0,    5,    8,    7,   17,    8,    3,   17,    1]
[   1, 7817,   19,    3,    2,    6,    0,    5,   24,    0]
[   1,   21, 6749,   27,   35,   11,   18,   46,   73,    9]
[   1,    5,   69, 6880,    1,   83,    5,   16,   50,   31]
[   2,    8,   15,    2, 6698,    3,   19,    4,   18,   55]
[  21,    7,   15,   76,   20, 6077,   23,    4,   48,   22]
[  11,    5,   13,    0,   17,   32, 6787,    2,    9,    0]
[   2,    5,   28,    7,   22,    5,    1, 7178,    4,   41]
[  11,   14,   25,   48,    7,   58,   18,    5, 6618,   21]
[  12,    3,    6,   20,   38,   15,    1,   38,   11, 6814]
```

Actual

Labeled

Joint Label Noise Distribution Matrix P(given_label, true_label) of shape (10, 10)

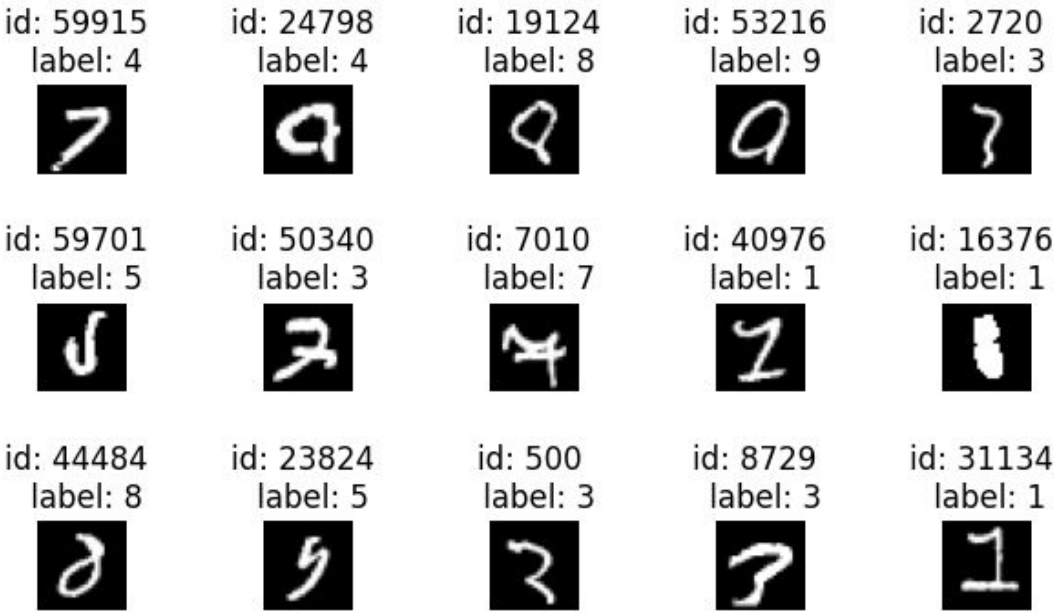
p(s,y)	y=0	y=1	y=2	y=3	y=4	y=5	y=6	y=7	y=8	y=9
s=0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
s=1	0.0	0.11	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
s=2	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
s=3	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0
s=4	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0
s=5	0.0	0.0	0.0	0.0	0.0	0.09	0.0	0.0	0.0	0.0
s=6	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0
s=7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0
s=8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.09	0.0
s=9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1

Trace(matrix) = 0.98

Results

143 out of 70000 labels identified as bad.

Many are mislabeled or questionable, some are correct and false positives.



Next Steps

- Increase cross validation from 3 folds to 10
- See difference in model accuracy without bad labeled data
- See how well it works with data that has uneven class distribution