CS 5350/6350: Machine Learning Fall 2022

Homework 1

Handed out: 6 Sep, 2022
Due date: 11:59pm, 23 Sep, 2022

1 Paper Problems

1.

(a) i. A larger hypothesis space implies higher sample complexity. Given that L_2 has a smaller hypothesis space, I prefer L_2.

ii. As the $|H|$ term in the inequality gets larger, m increases. Although L_1 and L_2 are both consistent with their respective training examples, the hypothesis space of L_2 is less complex. Occam's Razor states that, ceteris paribus, the simplest explanation is the best one. I think of it this way: if a hypothesis space is more complex, a learned function from that space has more degrees of freedom along which the function can be incorrect when generalized to new data.

(b)

$$|H| = 3^{10}; \delta = 0.05; \varepsilon = 0.10$$
$$m > \frac{1}{0.10}\left(\ln\left(3^{10}\right) + \ln\left(\frac{1}{0.05}\right)\right)$$
$$m > 140$$

2. The key to this proof is remembering that the sum of the incorrect-example-weights and the sum of the correct-example-weights add to 1. We can substitute the sum of

these two sums in place of 1 on the 5th line.

$$
\begin{aligned}
\epsilon_i &= \frac{1}{2} - \frac{1}{2}\left(\sum_{y_i=h(x_i)} D_t\left(i\right) y_i h\left(x_i\right) + \sum_{y_i\neq h(x_i)} D_t\left(i\right) y_i h\left(x_i\right)\right)\\
&= \frac{1}{2} - \frac{1}{2}\left(\sum_{y_i=h(x_i)} D_t\left(i\right)^* 1 + \sum_{y_i\neq h(x_i)} D_t\left(i\right)^* - 1\right)\\
&= \frac{1}{2} - \frac{1}{2}\left(\sum_{y_i=h(x)_i} D_t\left(i\right) - \sum_{y_i\neq h(x_i)} D_t\left(i\right)\right)\\
&= \frac{1}{2}\left(1 - \sum_{y_i=h(x)_i} D_t\left(i\right) + \sum_{y_i\neq h(x_i)} D_t\left(i\right)\right)\\
&= \frac{1}{2}\left(\sum_{y_i=h(x)_i} D_t\left(i\right) + \sum_{y_i\neq h(x_i)} D_t\left(i\right) - \sum_{y_i=h(x)_i} D_t\left(i\right) + \sum_{y_i\neq h(x_i)} D_t\left(i\right)\right)\\
&= \frac{1}{2}\left(2\sum_{y_i\neq h(x_i)} D_t\left(i\right)\right)\\
&= \sum_{y_i\neq h(x_i)} D_t\left(i\right)
\end{aligned}
$$

3.

(a) y = 1 if

$$X_1 - X_2 + X_3 \geq 2$$

(b) y = 1 if

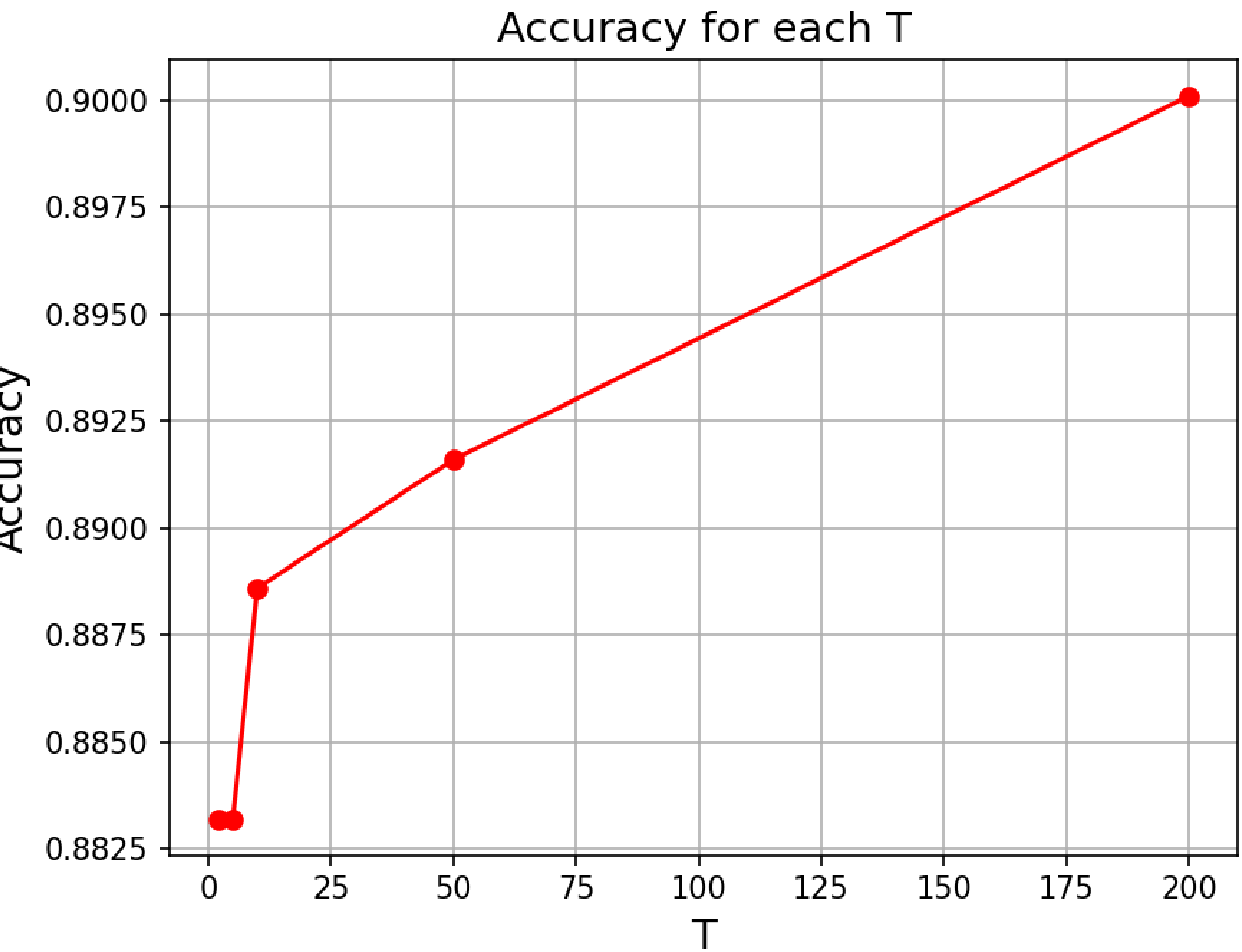
$$(1 - X_1) + (1 - X_2) + (1 - X_3) \geq 1$$

2 Decision Tree Practice

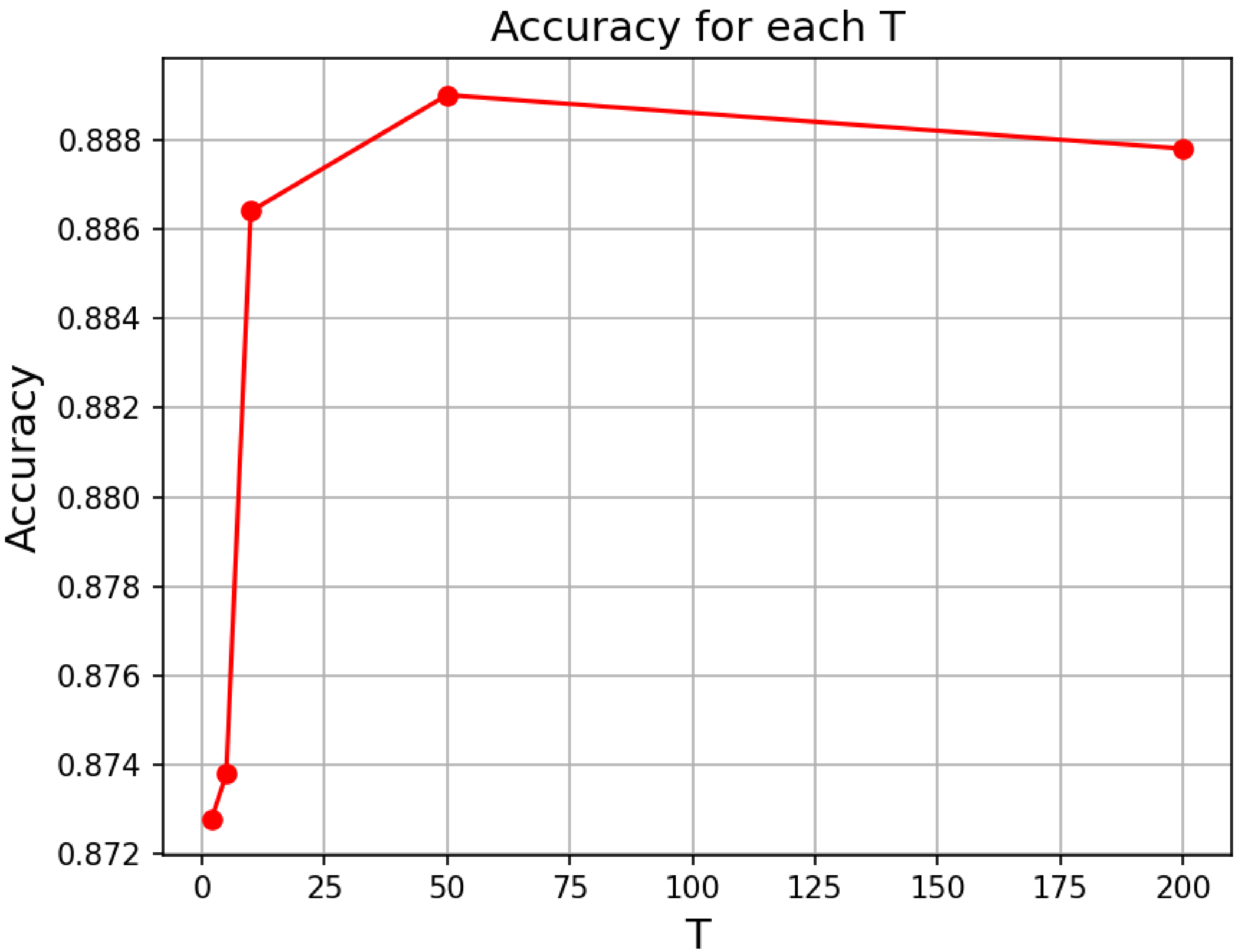
1. Link to repository: https://github.com/MattMyers204453/CS-6350.git

2.

(a) My conclusion is that Adaboost, in this case, is superior for large values of T. T = 500 was too slow to test.


Accuracy for each T
Accuracy
0.9000
0.8975
0.8950
0.8925
0.8900
0.8875
0.8850
0.8825
0
25
50
75
100
125
150
175
200
T

(b) My bagging algorithm seemed to stagnate past T = 100. My test sampled 1000 examples.


Accuracy for each T
Accuracy
T
0.888
0.886
0.884
0.882
0.880
0.878
0.876
0.874
0.872
0
25
50
75
100
125
150
175
200

3.

4. My Gradient Descent Algorithm found a solution with error very close to 0 with r = 0.5 and T = 100.