hw3: machine teaching for OLS

February 22, 2023

We now consider a teacher and a student.

- Given any dataset $D = (x_1, y_1) \dots (x_n, y_n)$, The student runs OLS to learn $(\hat{a}, \hat{b})$.
- The teacher knows the above student algorithm. The teacher can only influence the student with the teaching set D. The teacher wants to make sure the student learns a target model (a, b).

For this homework, we will again let the target model be $(a = 2, b = -3)$ so you can relate to hw2.

1 Q1: Synthetic Teacher

A synthetic teacher can create any (x_i, y_i) pairs it wants to include in D. We require that teaching is exact: $(\hat{a}, \hat{b}) = (a, b)$.

1. Create by hand a teaching set D_1 with $n = 2$
2. Create by hand another teaching set D_2 with $n = 3$, with the requirement that at least one pair is not on the $y = ax + b$ line.

For both D_1 and D_2, you should simulate the student to check for successful teaching, i.e. run OLS on your D_1 and D_2 and verify $(\hat{a}, \hat{b}) = (a, b)$.

Hand in: plot your D_1, D_2 together with the line $y = ax + b$.

2 Q2: Pool-based Teacher

A pool-based teacher cannot create arbitrary data points. Instead, the teacher is given a "pool" of data points $P = \{(x_1, y_1) \dots (x_N, y_N)\}$, and the teacher must select pairs from the pool to form its teaching set D. Repeated selection of the same pair is allowed (i.e. D is a multiset). Exact teaching is in general infeasible in pool-based teaching. Instead, the teacher aims to approximately teach the student the target model.

One simple but inefficient teaching algorithm is the following: fix n to be a small number like 2. Enumerate all possible D of size n with labeled items selected from the pool P. This is similar to N-choose-n except that we allow repeats. Run the student on each D to produce $(\hat{a}, \hat{b})$. Finally, pick the D that drives the student as close to the target (a, b) as possible. There are many ways to define closeness. For simplicity, we will use squared difference in parameters:

$$\ell(D) := (\hat{a} - a)^2 + (\hat{b} - b)^2.$$

Implement this algorithm. We give you the pool P in hw3pool.txt where the first column is x_i and the second column is y_i. Hand in:

1. Simply give the whole P to the student. Show $(\hat{a}, \hat{b})$. Show the value of $\ell(P)$.

2. Select the best data set D_3 with $n = 2$. Plot P, overlay D_3 (e.g. use different color). Show $(\hat{a}, \hat{b})$. Show the value of $\ell(D_3)$. Show the run time in seconds.

3. Select the best data set D_4 with $n = 3$. Plot P, overlay D_4 (e.g. use different color). Show $(\hat{a}, \hat{b})$. Show the value of $\ell(D_4)$. Show the run time in seconds.