AIL 7310: Machine Learning for Economics
AY 2025-26 Semester I, Assignment 4

1. **Double-LASSO:** Use the dataset titled 'sim_health.csv' to estimate the causal effect of a health program (binary participation D) on a continuous health outcome Y. The dataset contains many pre-treatment covariates (demographics, clinical measures, behaviors, environment, prior utilization, labs), so this is a high-dimensional problem where double/debiased lasso (DML) is appropriate for valid inference on a low-dimensional parameter (the average effect of participation). Make sure to implement a cross-fitting approach.

2. **Causal Forest:** In question 1 above, we used double-LASSO to get the Average Treatment Effect. Now we want to get Conditional Average Treatment Effect. Here, the goal is to evaluate not only whether the program works (the average treatment effect, ATE), but also for whom it works best (CATE).

   (i) Estimate the Average Treatment Effect here and compare with the Double LASSO.

   (ii) Estimate Conditional Average Treatment Effect and plot it with respect to
   a) Region
   b) Household Income