

Modelo descriptivo

Anzony Quispe

En esta entrega del seminario les comentaré acerca del modelo matemático que seguiré en la tesis. Este modelo está basado en el modelo de Automatic Debiased Machine Learning (Chernozhukov et al., 2022). Esta es una extensión del trabajo primario de Chernozhukov et al. (2014). Intentaré explicar cómo el método consigue estimar de manera consistente un parámetro de dimensión finita ante la presencia de parámetros de alta dimensión siguiendo un Partial Linear Modelo tal como lo hace el autor originalmente.

Los métodos de estimación de Machine Learning son particularmente precisos al predecir en un contexto de alta dimensión, es decir, cuando los parámetros son muchos, incluso más que las observaciones en el *dataset*. Esta propiedad viene del hecho que utilizan procesos de regularización que hace posible que el aprendizaje sea posible en los modelos. La penalización reduce varianza y añade sesgo a cambio de *overfitting*. Estos últimos conducen a una estimación sesgada del parámetro a estimar. Sin embargo, los autores encuentran que este sesgo puede ser eliminado si utilizamos momentos Neyman-Orthogonal y cross-fitting. El primero estima el modelo ortogonalizando el tratamiento con el resto de las variables y quedándose únicamente con el residuo de la regresión. Un proceso muy parecido al que seguimos en Frisch-Waugh-Lovell. Para este proceso utilizamos un modelo ML y lo ajustamos con una porción auxiliar de la data. De esta parte proviene el nombre de double machine learning, puesto que estamos estimando dos modelos, uno para la regresión principal y otro en la ortogonalización. Este proceso extra permite eliminar el sesgo generado por la regularización del modelo original, en otras

palabras, logra que el parámetro sea menos sensible al resto de parámetros distractores. Sin embargo, aún después de eliminar este sesgo queda por resolver el sesgo inducido por overfitting. Para esto necesitamos asumir que las observaciones son independientes y que el modelo ML principal y auxiliar fue ajustado utilizando únicamente las observaciones auxiliares. Por lo tanto, cuando condicionamos en el sample auxiliar, este último término tiende a 0 en probabilidad. Con ambos modelos ajustados, usamos la data principal y obtenemos una estimación del parámetro, luego repetimos el proceso solo intercambiando la data y obtenemos otra estimación del parámetro, finalmente, obtenemos el promedio de ambos parámetros que sí es consistente.

Un paso hacia adelante en la literatura es que la ortogonalización de Neyman permite visualizar el término que quita el sesgo de parametrización (debiasing term). Utilizando el teorema de representación de Riez, se obtiene el resultado de quitar el sesgo de parametrización (Riez Representer). La literatura ha intentado caracterizar este término para luego estimarlo; sin embargo, este se puede estimar directamente. Se propone estimar este término con cualquier modelo de Machine Learning tomando en cuenta que el término proviene de un proceso de minimización del error entre el momento del tratamiento y el debiasing término, que a su vez es el error entre el sesgo de regularización verdadero y el estimado.

Referencias

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters.

Chernozhukov, V., Newey, W. K., & Singh, R. (2022). Automatic debiased machine learning of causal and structural effects. *Econometrica*, 90(3), 967-1027.