

Prediction challenge notes. App - ML 2020

Juan Ignacio

Lucas Mation

Expected MSE

Questions

- What is the correct protocol for dealing with missing values in some X? In reg-world to solve the problem of X having missings for some observations, one creates a new dummy, X_{mi} , to indicate missing values and then replaces the missing values in X by any arbitrary number (eg: zero, or the average of X) and then include both X and X_{mi} in the regression. We did something similar for the missings in our variables, replacing them by the average of X. However, we are worried about the ML algorithm not picking up the missing indicator dummy (X_{mi}), making the procedure sensitive to the imputed value.
- How to incorporate the fact that Y is a count data into the ML estimation? In regression we would use a poisson link function. But we did not incorporate that into our procedure.

Lessons learnt

- We spent 95% of the time to understand and prepare the data
- Prediction is hard!