

Predicition challenge notes. App - ML 2020

Juan Ignacio

Lucas Mation

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.

Lessons learnt

Expected MSE