A study of MLP for the imputation of the "Attained Level of Education" in Base Register of Individuals

Fabrizio De Fausti, Marco Di Zio, Romina Filippini, Simona Toti[1], Diego Zardetto[2]

Abstract

The Attained Level of Education (ALE) of the Permanent Italian Census relies on a high amount of administrative information. Nevertheless, it is needed to resort to sample survey data to cope with delay of information and coverage problems. Due to the complexity and heterogeneity of the available information, the solution of the problem with standard statistical methods needs the construction of different imputation models with a strong effort in terms of human intervention. We study the use of a multilayer perceptron (MLP) model to make the process more automatic, i.e., less costly in terms of human resources, and possibly more accurate in terms of estimates. Since a relevant quality aspect is the ability of the imputation process to provide a good estimate of the ALE frequency distribution, sampling weights are used in both classic statistical techniques and in the context of machine learning.

Keywords: Register-based statistics, imputation, data integration, sampling weights, machine learning.

---

[1] Istat, e-mail: defausti@istat.it , dizio@istat.it, filippini@istat.it, toti@istat.it.
[2] World Bank, dzardetto@worldbank.org, and Istat, zardetto@istat.it.