Proposed Framework for complete analysis when teaching Regression in Supervised Machine Learning

Charles Alba

1The Pennsylvania State University, University Park, USA
cja5553@psu.edu

INTRODUCTION

Teaching Regression in Machine Learning for analytics could be challenging for instructors and students alike. This could be particularly challenging for primary and secondary school teachers and students alike when attempting to learn about Regression in analytics for the first time. Unlike teaching in most conventional subjects where the concepts are simply explained, then examples are given before students apply the taught concepts, regression in Machine Learning has many intertwined concepts that are crucial to each other. For instance, most instructors rightfully spend huge chunks of separate lessons separately teaching assumptions to regression, the different types of regression and its applications, the test statistics behind each of them, etc.

The detrimental effects resulting from the huge learning gap between “novices” – defined as those who are “new to the skill” (Davis, Horn and Sherin, 2013) [1] – and “experts” – defined as those who have training in the skill (Chi, Feltovich, and Glaser 1981; Chi, Glaser, and Rees 1982) [2][3] – could be applied to teaching Regression in Analytics to Primary and Secondary School kids. Where these kids are expected to learn seemingly advanced concepts for the first time by instructors who have experience with relevant analytical skills, issues could be foreseen by both students and teachers. Where teaching regression in supervised machine learning may also incorporate some form of programming, primary and secondary students may find learning such topics daunting (Perkins, Hancock, et al, 1986) [4]. Likewise, piecing together a curriculum and instructing a subject that is uncommon in Primary and secondary school levels for the first time may pose uncertainty to teachers (Carlsen, 1987) [5]. Such could be further discouraging when the visible confusion of students does trickle down to the teacher, causing them to be de-motivated (Civil, 1992) [6].

The core of Machine Learning requires students to utilize a raw data set and perform a complete analysis of such. Therefore, it is understandable that students and instructors alike would have a difficult time piecing together the taught concepts to perform a complete data analysis. Even the most organized students could potentially miss out on some trivial yet vital steps in piecing together a complete regression analysis.

Hence, we hope to integrate 2 concepts when teaching Regression in Supervised Machine Learning to Primary and Secondary school students: repetition and systematic instruction. Repetition, while despised by some educators, could be necessary to achieve proficiency (Wells and Hagman, 1989; Thalheimer, 2006) [7][8]. Learning analytics at such a young stage is certainly no exception. Similarly, when introducing seemingly complicated concepts to young learners for the first time, systematic and explicit instruction is found to be effective and useful (Paulsen, et al, 2017) [9]. This effectiveness is further when teaching mathematics-based subjects in “improving a student’s ability to solve problems” (Paulsen, et al, 2017) [9].

Hence, one suggested instructional hack to mitigate this issue is the introduce and consistently highlight a framework towards performing a complete Regression Analysis. This framework details the steps
towards regression in Supervised Machine Learning. This could be reiterated throughout the course regardless if the student understands what each step entails. The idea is to ensure the students could learn Regression in supervised Machine Learning in a structured and systematic manner that would allow them to combine the building blocks with ease when it comes time to perform a complete regression analysis.

**Framework**

| Step | Task |
|------|------|
| 1.   | Tidy the raw data |
| 2.   | Identify the Data Type |
| 3.   | Deal with the missing data, if any |
| 4.   | Based on (2), identify the right type of regression Analysis |
| 5.   | Test the necessary assumptions |
| 6.   | Make modifications to (4) if (5) fails |
| 7.   | Perform regression analysis |
| 8.   | Complement (7)** |
| 9.   | Interpret the results with context to the question |

**Details**

Step (1): An often-overlooked step, this could potentially save the student so much time in terms of coding. This illustrates why such a framework is needed, because students would forget these trivial yet crucial steps.

Step (2): Involves identifying if the Data is categorical vs Quantitative as well as if the Response vs Predictor Variable

Step (3): Involves the utilization of techniques such as Data Imputation if the given raw data set has cases of empty data sets.

Step (4): Based on the data type, when students have comprehensively learned the different types of regression analysis, they could identify the best regression analysis based on (2)

Step (5): Based on the relevant regression analysis identified at (4), students should ensure the assumptions are met before proceeding. This is yet another crucial step student tends to forget due to the inability to piece lessons in a structural manner.

Step (6): Involved making model diagnostics if assumptions are not met.

Step (7): Self-explanatory, involves the use of the relevant Test Statistic.

Step (8): Could involve the use of graphs for visualization purposes or techniques like stepwise regression.

Step (9): Self-explanatory
Discussions & Conclusions

One could foresee how introducing and consistently reiterating such a framework throughout a coursework involving regression in Supervised Machine Learning would ensure a student is able to connect the different steps which could be taught separately in large chunks. This is further pertinent given how at times an instructor may understandably not teach Regression structurally based on the steps in order to facilitate easier understanding across intertwined concepts. For example, an instructor may understandably choose to teach the assumptions first before delving into details of a regression technique, while some may not rightfully teach steps (1) to (2) because a student is expected to have learned such from elementary statistics while some may omit concepts like imputation if time permits, etc.

It could be suggested that instructors could utilize such a framework in the syllabus of the course or its introduction. Such would pave the way for conceptualizations in a student’s understanding as to what the coursework with regards to Regression in Supervised Machine Learning would entail. It would also allow students to think of the coursework systematically.

Instructors could continuously refer the students to the framework when moving on from a concept. This would allow students to picture which procedure of Regression is being taught and if possible, relate previously taught concepts to concepts that are about to be taught. Such also paves the way for reiteration of past concepts and the steps towards regression analysis, which has been proven to be helpful in education[10].

References

1. Davis, P. R., Horn, M. S., & Sherin, B. L. (2013). The Right Kind of Wrong: A “Knowledge in Pieces” Approach to Science Learning in Museums. *Curator: The Museum Journal, 56*(1), 31-46. https://doi.org/10.1111/cura.12005

2. Chi, M. T., Feltovich, P. J., & Glaser, R. (1981). Categorization and Representation of Physics Problems by Experts and Novices*. *Cognitive Science, 5*(2), 121-152. https://doi.org/10.1207/s15516709cog0502_2

3. Chi, M. T., Glaser, R., & Rees, E. (1982). Expertise in problem solving: Advances in the psychology of human intelligence. *Hillsdale, NJ: Erlbaum*, 1-75.

4. Perkins, D. N., Hancock, C., Hobbs, R., Martin, F., & Simmons, R. (1986). Conditions of learning in novice programmers. *Journal of Educational Computing Research, 2*(1), 37-55. https://doi.org/10.2190/GUJT-JCBJ-Q6QU-Q9PL

5. Carlsen, W. S. (1987). Why Do You Ask? The Effects of Science Teacher Subject-Matter Knowledge on Teacher Questioning and Classroom Discourse.

6. Civil, M. (1992). Prospective Elementary Teachers’ Thinking about Teaching Mathematics.

7. Wells, R., & Hagman, J. D. (1989). *Training procedures for enhancing reserve component learning, retention, and transfer* (Technical Report no. 860). Alexandria, VA: US Army Research Institute for the Behavioral and Social Sciences. (AD A217 450).

8. Civil, M. (1992). Prospective Elementary Teachers' Thinking about Teaching Mathematics.

9. Paulson, K., et al. (2017). High-quality mathematics instruction: What teachers should know. Retrieved from https://iris.peabody.vanderbilt.edu/module/math/

10. Fisher, Mercer. *How do teachers help children to learn? An analysis of teachers' interventions in computer-based activities*, Learning and Instruction, 1990(2), 339-355