Week 7 Assignment: Presenting and Interpreting Results of Predictive Models

Jonathan Ibifubara Pollyn

College of Science, Engineering and Technology, Grand Canyon University

DSC-530: Predictive Modeling

Filippo Posta

October 20, 2021

Part 1: Operational Tasks

The generation of the contingency table shown below

Predicted / Actual	0	1	Total
0	17665	7269	24934
1	751	24013	24764
Total	18416	31282	49698

by comparing the anticipated response values from Model 1 to the actual response values from the loans data driven test data set. TAN = 24,934, TAP = 24,764, and GT = 49698 would be true for any model, as these are observed values, not predicted ones. The remaining values vary according to the predictions made by the various models. The TAN = 24934 negative records in this scenario, out of a total of GT = 49698 data, implying that all negative Model, which labels all forecasts as negative, would have an accuracy of 0.

Thus, Model 1 does manage to narrowly outperform the baseline see the below calculation

The accuracy is obtained by Accuracy $= \frac{TN+TP}{GT} = \frac{17665+24013}{49698} = 84\%$

The error rate is obtained by Error Rate $= \frac{FN+FP}{GT} = \frac{751+24934}{49698} = 52\%$

The sensitivity is obtained by sensitivity $= \frac{TP}{TAP} = \frac{24013}{24764} = 97\%$

The specificity is obtained by specificity $= \frac{TN}{TAN} = \frac{17665}{24934} = 71\%$

The precision is obtained by precision $= \frac{TP}{TPP} = \frac{24013}{31282} = 77\%$

The F_1 is obtained by $F_1 = 2 * \frac{(TP/TPP)(\frac{TN}{TAN})}{(\frac{TP}{TPP})+(\frac{TN}{TAN})}$ =2* $\frac{(0.77)(0.71)}{(0.77)+(0.71)}$ = 74%

The F_2 is obtained by $F_2 = 5 * \frac{(TP/TPP)(\frac{TN}{TAN})}{4*(\frac{TP}{TPP})+(\frac{TN}{TAN})}$ =5* $\frac{(0.77)(0.71)}{5*(0.77)+(0.71)}$ = 60%

The $F_{0.5}$ is obtained by $F_{0.5} = 1.25 * \frac{(TP/TPP)(\frac{TN}{TAN})}{0.25*(\frac{TP}{TPP})+(\frac{TN}{TAN})}$ =1.25* $\frac{(0.77)(0.71)}{0.25*(0.77)+(0.71)}$ = 75%

PART 2: Mathematical and Statistical Basis

Compare the z model (ORM) given with the continuation ratio model (CRM) and the adjacent category model.

Guzman-Castillo, M., Brailsford, S., Luke, M., & Smith, H. (2015). A tutorial on selecting and interpreting predictive models for ordinal health-related outcomes. *Health Services and Outcomes Research Methodology*, *15*(3-4), 223–240. https://doi.org/10.1007/s10742-015-0140-6

> The authors stipulated that model for ordinal data are classified according to their handling of the proportional odds or parallel regression assumptions. The authors stated that parallel regression indicates the beta coefficients will be the same across simultaneous regressions; assuming comparable slopes across categories enables models to be interpreted consistently across categories, resulting in more economical models. The authors pointed out that ordinal regression, continuation ratio, and adjacent category models are all based on this assumption. However, the widespread agreement is that this assumption is overly strict and that the likelihood of all dependent variables in the model

having equal slope coefficients is extremely unlikely. As a result, the authors stated that various alternative models had been proposed in the statistical literature. According to the authors, the ordinal regression model (ORM), alternately referred to as the cumulative odds model or proportional odds model, was the first model created specifically for ordinal outcomes. The authors also add that the ORM is frequently expressed in terms of latent variables. The authors also stated that the continuation ratio model (CRM) is a subtype of the ordinal model. The categories represent levels, with the lowest preceding the second, the second preceding the third and the highest. They affirmed that it could be viewed as steps in a process that an individual can progress through. The procedure is defined by the fact that an individual must complete each stage. The authors then stated that the adjacent probability is essential in the adjacent category model (ACM): having a short LoS versus having a medium LoS or having a medium LoS versus the probability of having a long LoS.

Compare the models, including the generalized ordered logit model (GOLM), sequential, and multinomial logit model (MNLM).

Guzman-Castillo, M., Brailsford, S., Luke, M., & Smith, H. (2015). A tutorial on selecting and interpreting predictive models for ordinal health-related outcomes. *Health Services and Outcomes Research Methodology*, *15*(3-4), 223–240. https://doi.org/10.1007/s10742-015-0140-6

Regarding the generalized ordered logit model (GOLM), the authors specified that the model allows distinct slope coefficients for each binary regression. They emphasized that GOLM's equations are similar to those of the ORM. The asserted that GOLM preserves the ORM's nature by examining the impacts of a set of independent factors

simultaneously across consecutive dichotomizations of the outcome while allowing the slope coefficients to change between the categories. In the multinomial logit model (MNLM), the authors stated that the model extends the binary logistic regression model for handling polydomous outcomes. The authors also affirm that categories are no longer regarded as ordered, and the effects of the independent variables are permitted to vary by the result. The authors finally settled that the MNLM is based on independent irrelevant alternatives, meaning that the chances do not depend on other possible outcomes.

Evaluate the partial proportional odds model (PPOM), the partial continuation model (PCM), and the stereotype ordered regression model (SORM)

Guzman-Castillo, M., Brailsford, S., Luke, M., & Smith, H. (2015). A tutorial on selecting and interpreting predictive models for ordinal health-related outcomes. *Health Services and Outcomes Research Methodology*, *15*(3-4), 223–240. https://doi.org/10.1007/s10742-015-0140-6

The authors confirmed that parallel lines are constrained in the partial proportional odds model (PPOM) only when required. The GOLM equation is extended to include the parameters that were not limited and so violated the assumption. The PPOM was calculated using the gologit2 command with the auto option, imposing constraints on variables that contradict the parallel assumption. The authors also affirmed that the constraints had to be manually inserted, although the PCRM was estimated using seqlogit. The partial continuation model (PCRM) extends the CRM equation by including coefficients for variables that break the correspondence assumption. The two models PPOM and PCRM, contain beta coefficients that change across categories solely for variables that contradict the parallel assumption; as a result, the authors stated that the

parameter estimates for the other variables are identical. The authors also stipulated that the stereotype ordered regression model (SORM) can be conceived of as constraining a multinomial model's ordering. The model was introduced as a counter-example to the ORM's stringent parallel regression assumption. Although the model was initially developed for ordinal data derived from an assessment procedure the authors confirmed, it is not limited to this variable type. While the SORM interpretation of the changes may be comparable to the MNLM, when computed using STATA's default parameters, it compares the highest and lowest categories.

Part 3: Christian Worldview

Discuss human dignity and value from the Christian worldview when your behaviors are predicted by the computer. Is it okay to be treated by a bank that you are not trustworthy because you delayed several payments?

Life and dignity of the human person | usccb. (n.d.). https://www.usccb.org/beliefs-and-teachings/what-we-believe/catholic-social-teaching/life-and-dignity-of-the-human-person

The author referred to the biblical definition; dignity is completely connected to the biblical concept of splendor. The Bible employs the terms "God's glory," "His weightiness," "His importance," and "His significance" to characterize the source of all dignity. The author went on to state that only God is eternally valuable and significant intrinsically. We are a creature made of dust, according to the author, and the dust is insignificant. Still, we become significant when God chooses it and makes it into a human being, breaths life into it, and declares, "This creature is formed in my image." The author stipulates that God invests transitory creatures with eternal importance. We

possess nothing in us that would compel God to see us as eternally significant. The author stated that we have lasting significance and value because God bestows them to us. God offers it to every human being; that is why the great commandment in the Bible addresses our connection with God and our relationship with one another. "Thou shalt love the Lord your God with all your heart, soul, and might...and your neighbor as much as you love yourself," because God has endowed every human creation with worth. It could be a thing of false prediction if a computer is a factor that decides the dignity of humans. God made us in his image, which he says we are worthy of, but the computer decides it based on our past and search history. As Christians, the Bible says in 1 John 1:9, "If we confess our sins, he is faithful and just and will forgive us our sins and purify us from all unrighteousness." the computer does not forgive but keep every sin we committed as a factor to make a decision.

References List

Guzman-Castillo, M., Brailsford, S., Luke, M., & Smith, H. (2015). A tutorial on selecting and interpreting predictive models for ordinal health-related outcomes. *Health Services and Outcomes Research Methodology*, *15*(3-4), 223–240. https://doi.org/10.1007/s10742-015-0140-6

Larose, C. D., Larose, D. T., & Larose, Chantal D., Author. (2019). *Data science using python and r*. John Wiley & Sons,inc,.

Life and dignity of the human person | usccb. (n.d.). https://www.usccb.org/beliefs-and-teachings/what-we-believe/catholic-social-teaching/life-and-dignity-of-the-human-person