Computerized Adaptive Test

1. **Problem statement:** *. Create a Neural Network to analyse response accuracy per question and modify the weights to hence obtain a new weighting for the question set, which will better suit the current demographic.*

2. **Description:** *Computerised Adaptive Test, consists of questions weighted as per difficulty and presented to the examinee one at a time. The response accuracy determines difficulty of subsequent questions, which results in a balanced test for people of all calibres. The perception of difficulty may vary depending on the demographic attempting the test, which would lead to examinees experiencing an unbalanced question distribution*
3. **Overview and Objectives:** *A software that tests examinees as per their caliber which uses Machine learning Algorithms to train our model and implement probabilistic approaches for updating weights per question.*

• To read a set of responses per question, for the entire set, as input.
• To train the model on said training-input and assign difficulty weights per question.
• To read a set of questions from the set and their response, correct or wrong, as test-input.
• To record responses and add it to your training dataset.
• To output the final weights assigned to the question set after processing all inputs.
• To device an optimization strategy for the model and the weights it assigns.
• To make the model self-reliant and adaptive to diverse response sets.

4. **Constraints:**

• The training-input set will be provided as a [csv] file of the following nature:
• The rows contain responses of a student for each question.
• The test-input should be used as the last 100 rows(students: 900-999) of the training set, and any 10 columns (questions: Qx). • Final test input shall be applied on your system when it is submitted.
• You are allowed to use any Deep Learning approach, publicly available resources, and additional training data, but the final output should be on the provided list.
• The output shall be in the form of an auto-generated spreadsheet having a column for question number and another for weight assigned to said question.

5. **Rasch model:** *In the Rasch model, the probability of a specified response (e.g. right/wrong answer) is modeled as a function of person and item parameters. Specifically, in the original Rasch model, the probability of a correct response is modeled as a logistic function of the difference between the person and item parameter. The mathematical form of the model is provided later in this article. In most contexts, the parameters of the model characterize the proficiency of the respondents and the difficulty of the items as locations on a continuous latent variable. For example, in educational tests, item parameters represent the difficulty of items while person parameters represent the ability or attainment level of people who are assessed. The higher a person's ability relative to the difficulty of an item, the higher the probability of a correct response on that item. When a person's location on the latent trait is equal to the difficulty of the item, there is by definition a 0.5 probability of a correct response in the Rasch model.*

Scaling

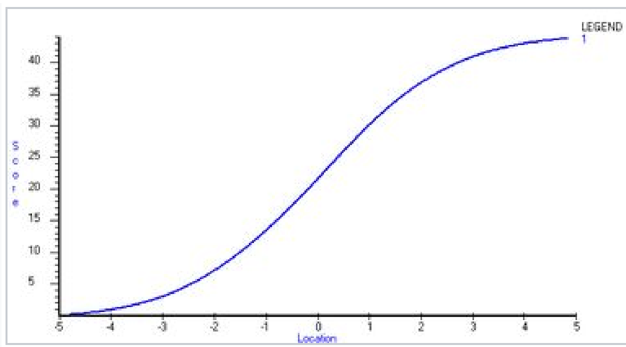


Figure 1: Test characteristic curve showing the relationship between total score on a test and person location estimate

Interpreting scale locations

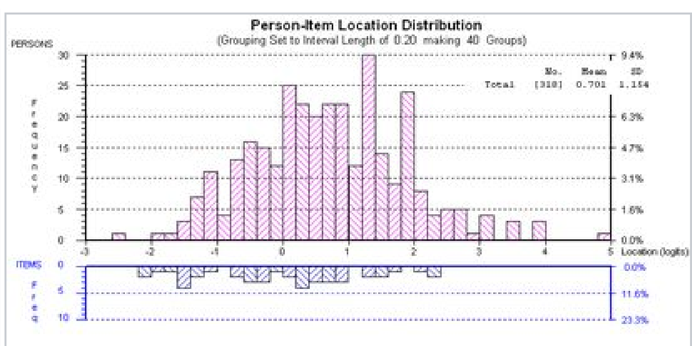


Figure 2: Graph showing histograms of person distribution (top) and item distribution (bottom) on a scale

1. *When all test-takers have an opportunity to attempt all items on a single test, each total score on the test maps to a unique estimate of ability and the greater the total, the greater the ability estimate. Total scores do not have a linear relationship with ability estimates. Rather, the relationship is non-linear as shown in Figure 1. The total score is shown on the vertical axis, while the corresponding person location estimate is shown on the horizontal axis. For the particular test on which the test characteristic curve (TCC) shown in Figure 1 is based, the relationship is approximately linear throughout the range of total scores from about 10 to 33. The shape of the TCC is generally somewhat sigmoid as in this example. However, the precise relationship between total scores and person location estimates depends on the distribution of items on the test. The TCC is steeper in ranges on the continuum in which there are a number of items, such as in the range on either side of 0 in Figures 1 and 2.*
2. *In applying the Rasch model, item locations are often scaled first, based on methods such as those described below. This part of the process of scaling is often referred to as item calibration. In educational tests, the smaller the proportion of correct responses, the higher the difficulty of an item and hence the higher the item's scale location. Once item locations are scaled, the person locations are measured on the scale. As a result, person and item locations are estimated on a single scale as shown in Figure 2.*

ICC plots

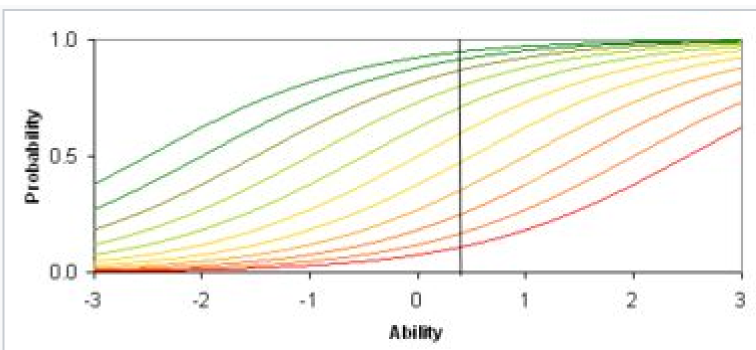
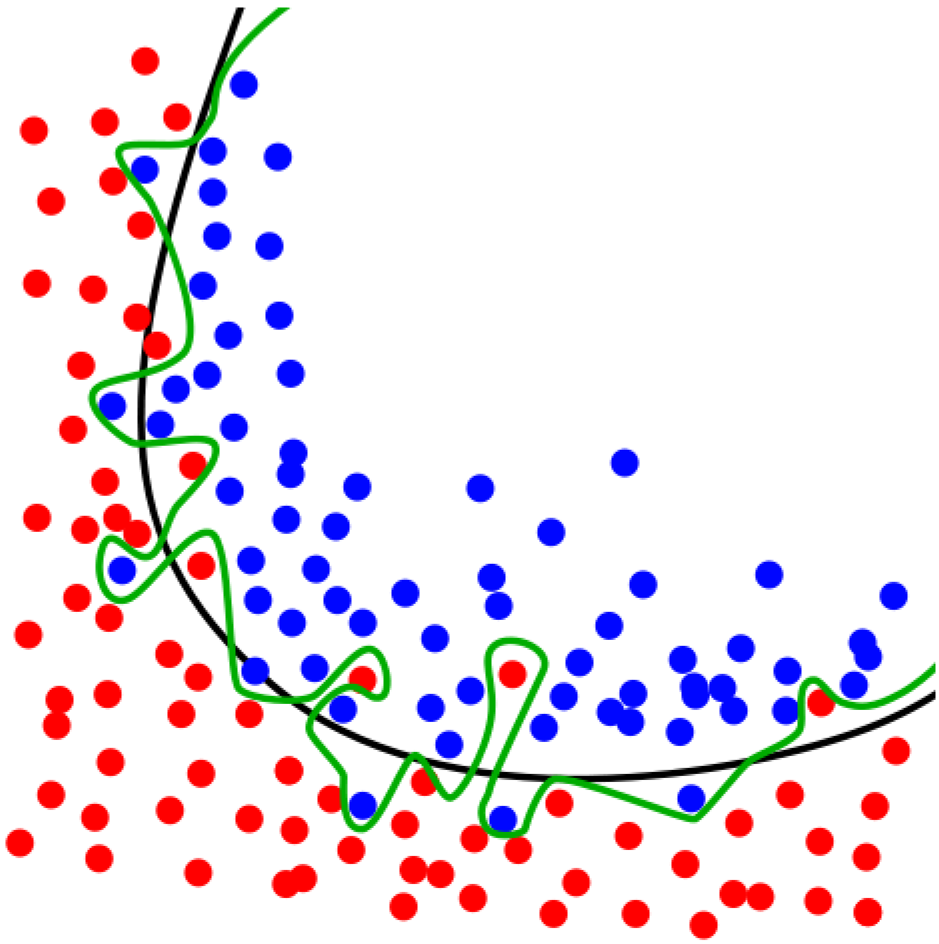


Figure 3: ICCs for a number of items. ICCs are coloured to highlight the change in the probability of a successful response for a person with ability location at the vertical line. The person is likely to respond correctly to the easiest items (with locations to the left and higher curves) and unlikely to respond correctly to difficult items (locations to the right and lower curves).

6. **Classification algorithm(knn):** *Classifying outcomes into 1’s and 0’s is a classic problem in data science.We use kNN is a non-parametric algorithm that is then free from assumptions about the relationship between the target and feature.*

Our extension of KNN with automatic feature weighting, multi-class prediction, and probabilistic inference, enhance prediction accuracy significantly while remaining efficient, intuitive and flexible. This general framework can also be applied to similar classification problems involving heterogeneous datasets.

0 ***1***



#REFERENCES#

Barrada JR, Abad FJ, Olea J (2014). “Optimal Number of Strata for the Stratified Methods in Computerized Adaptive Testing.” The Spanish Journal of Psychology, 17. ISSN 1138-7416. doi:10.1017/sjp.2014.50. URL http://www.journals.cambridge.org/

Barrada JR, Mazuela P, Olea J (2006). “Maximum Information Stratification Method for Controlling Item Exposure in Computerized Adaptive Testing.” Psicothema, 18(1), 156– 159. ISSN 02149915. URL http://156.35.33.98/reunido/index.php/PST/article/ view/8411.

Barrada JR, Olea J, Ponsada V, Abad FJ, Ponsoda V, Abad FJ (2009). “Test Overlap Rate and Item Exposure Rate as Indicators of Test Security in CATs.” In Proceedings of the 2009 {GMAC} Conference on Computerized Adaptive Testing. URL http://iacat.org/ sites/default/files/biblio/cat09barrada.pdf.

Barrada JR, Olea J, Ponsoda V (2007). “Methods for Restricting Maximum Exposure Rate in Computerized Adaptative Testing.” Methodology, 3(1), 14–23. ISSN 16141881. doi: 10.1027/1614-2241.3.1.14. URL http://psycnet.apa.org/journals/med/3/1/14/.

Barrada JR, Olea J, Ponsoda V, Abad FJ (2010). “A Method for the Comparison of Item Selection Rules in Computerized Adaptive Testing.” Applied Psychological Measurement, 34(6), 438–452. ISSN 0146-6216. doi:10.1177/0146621610370152. URL http://apm. sagepub.com/cgi/doi/10.1177/0146621610370152.

Barrada JR, Olea J, Ponsoda V (2007). “Methods for Restricting Maximum Exposure Rate in Computerized Adaptative Testing.” Methodology, 3(1), 14–23. ISSN 16141881. doi: 10.1027/1614-2241.3.1.14. URL http://psycnet.apa.org/journals/med/3/1/14/.

Chalmers RP (2016). “Generating Adaptive and Non-Adaptive Test Interfaces for Multidimensional Item Response Theory Applications.” Journal of Statistical Software, 71(5). ISSN 1548-7660. doi:10.18637/jss.v071.i05. URL http://www.jstatsoft.org/v71/i05/.