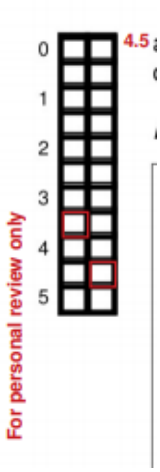
Notes on Business Analytics Exam Grading and the Grade Review Process

General



- On the cover page of the review process, you will see (a) your personal details, (b) your final exam credits (points) and grade (see below) and (c) possibly some annotations for the grading. If you participated in the Analytics Cup, the grade seen in (b) is the grade for the exam only (i.e. without the AC bonus). However, the grade you see in TUMonline will include the AC bonus, so it might be different from the one listed on the cover page. We have sparingly used the annotations in (c). (Not all correctors used the annotation system, so also check any handwritten notes.) If you make a request for regrading (see below), we will use the annotations (c) to explain our decision in regrading.
- Each problem in your exam has been graded twice. The **red** comments were written by the first corrector. The second corrector has checked the work of the first corrector and may have made additional comments in **green**. Both correctors have marked your credits for that exercise by checkmarks in the margin of the page. The left column of assigned credits corresponds to the first corrector, the right column corresponds to the second corrector. The final grade for a (sub)exercise is also indicated by a red number in the margin.
- In cases of disagreement between the correctors, the credits assigned by the second corrector (**green**) are binding. Your final exam grade is thus based only on the sum of credits marked in green / in the right column. In the example on the top right, the first corrector gave 3.5 points, the second corrector overrode this decision and gave 4.5 points, which are counted toward your total points on the cover page.

- **If you used the pre-printed individual paper,** rather than the exam booklet, then unfortunately TUMexam cannot correctly display your answers alongside the grading. In this case, you can only see the reviewer's handwritten comments (if any), the annotations on the cover page, and the number in the margin indicating your final grade, see the example on the bottom right. *We apologize for the inconvenience but please understand that we cannot provide a better exam review quality for individual papers which already caused significant overhead in grading. The alternative would be banning individual sheets completely.*
- If you believe that an exercise was not graded correctly, you can make a request for regrading of the specific (sub)problem via the TUMexam Grade Review process. In this case, a TA will review your request and grade the exercise again. **Make sure you read and follow all information in this document before making a request. Failure to do so will result in your request being denied without further explanations.** Note that, as a consequence of any request, we will completely regrade the entire (sub)problem, which may also result in the *loss of* points, not only improvements.
- To make a request: (1) carefully read the information about the grading scheme and common mistakes for the problem, which are provided below. (2) If you still believe the grading is incorrect, use the given space (500 characters) to make a *clear and concise* argument about the mistake in grading, simply writing "I believe I deserve more points." is not enough.

Problem 1 – Regression

- There were **15 points** to be reached in total.

a) **1 point** was awarded for correct dependent and independent variables. **0.5 points** was awarded for Identifying the variable as nominal/categorical. **0.5 points** were awarded for a factorization with three dummy variables. Note that processing it with four variables would yield multicollinearity and is thus incorrect.

b) To reach full credit of **2 points**, you were expected to provide the formula for the likelihood function L, including a correct expression of the probability and using the variable names (including snowfall as "y").
Partial credit was given for the general formula from the script, for partially correct or incomplete attempts to incorporate the variable names, or for writing down parts of the likelihood function (e.g. the sigmoid function).

c) **1 point** was awarded for identifying the significant variables and an explanation with t-test or p-values. Answers like "look at the stars" are not a sufficient explanation. Please note that variables that are significant at 1% or 0.1% are also significant at 5%.
1.5 points were awarded to the interpretation of the intercepts. You were expected to refer to the odds, meaning you were expected to calculate $\exp(\beta_0)$ and to interpret this value as odds when all independent variables are zero. Partial credit was given accordingly. Note that the intercept indeed explains a shift of the curve, but this interpretation does not refer to the odds.
1.5 points were awarded to the interpretation of one variable coefficient. You were expected to refer to the odds, meaning we expected to calculate $\exp(\beta_i)$ and to interpret this value as the factor increase/decrease in odds when the independent variable increases by one unit. Partial credit was given accordingly. Statements like "snowfall decreases" or "likelihood of snowfall increases by factor…" are not correct, since the odds are increasing/decreasing by certain factors.

d) **1 point** was awarded for interpreting the McFadden R^2 correctly. Note that a value between 0.2 and 0.4 is acceptable, but not a very good fit. "Good fit" was thus only awarded with partial credit.
2 points were awarded for a conceptual comparison of McFadden R^2 and OLS R^2. You were expected to explain that the McFadden R^2 compares the loglikelihood of full and null model, while the OLS R^2 measures the proportion of explained variance. Partial credit was awarded accordingly. Note that answers like "it measures the fit" or "R^2 of OLS is usually higher" are unspecific and do not explain conceptual differences.

e) **1 point** was awarded for statements about the McFadden R^2 and the significance. **3 points** were awarded for identifying multicollinearity and suggesting the Variance Inflation Factor. Note that the question explicitly referred to the revision of the model. Endogeneity might occur, but this is not evident from comparing the two models. Similarly, the output gives no indication for autocorrelation or heteroscedasticity. Examining correlations can help to detect pairwise multicollinearity and was awarded with partial credit.

Problem 2 – Data Preparation

- There were **10 points** to be reached in total.
- For each issue in the presented data set, you were awarded up to **2 points**. (There were more than 5 issues present, but you could not reach more than 10 points in total, even if you correctly described 6 or 7 issues.)
- To reach 2 points for a given issue, you had to (a) correctly identify and describe the issue, and (b) describe a possible solution to overcome it. Partial credit was given for correct issues with missing and/or incorrect explanations or solutions.
- If you listed the same issue twice (e.g. same issue in two different columns), points were only given once.
- The following answers were common but incorrect:
  - Some students confused the concepts of *statistical inference* (where you aim to fit and interpret model coefficients in order to understand the data generating process of a fixed data set) and *prediction* tasks (where you train a machine learning model in order to make good predictions on new, unlabeled instances). In the latter, rebalancing/resampling very unbalanced dependent variables can be an effective way to create models that generalize well to unseen data. However, this method is **not applicable** to statistical inference, as it changed the distribution of the underlying data set in which you are interested.
  - Some students suggested discretizing (e.g. binning) interval or ratio scaled (e.g. numeric) variables. This is not required, and indeed counterproductive, when applying methods that can natively handle numeric input data (i.e. regression models like here). This step would only be necessary for methods that inherently cannot retain numeric information and rely on discrete inputs, e.g. Naïve Bayes.
  - Nominal *features* (e.g. `character` type) do not need to be explicitly One-Hot-Encoded/factorized in R. (This is only required for the dependent variable.) If you mentioned this as an issue for a *feature* rather than the *label*, you were nevertheless awarded 1 point.
  - Correlation values of 70% - 80% are perfectly fine and do not cause concern about multicollinearity. (In the dataset, there were also pairs of columns with a correlation of >95%.)

Problem 3 – Identify Models

- There were **10 points** to be reached in total.
- For each plot/model you could reach **2.5** points. 1 point was given for each correct match of plot to model, 1.5 points were given for the explanation of the plot, and how you identified the match.
- If your match was incorrect, your explanation may nevertheless earn partial credit.
- Many students failed to answer (b), i.e. describe what exactly was seen in the plot. This, combined with imprecise/unclear language in (c) may often have resulted in loss of points for (c) even if you might have "meant" the right thing, but this was not clear to us from your answer.

Problem 4 – Evaluation

- There were **18 points** to be reached in total.

a) **2 points** were awarded for calculating accuracies for both classifiers. If you provided the correct formula but miscalculated both accuracies, you still received 0.5 points.

b) **2*2 = 4 points** were awarded for calculating the two metrics (similar to a.)). **0.5 points** were awarded for an interpretation referring to the capabilities of the two softwares. Note that it was not sufficient to state "Software A has a higher recall", since this is evident from the numbers.

c) **2 points** were awarded for identifying the correct metric, calculating them for both classifiers, and making a correct recommendation. Partial credit was awarded accordingly (e.g. recommending the correct software with a false argument). Note that a "high share of emails classified as spam is indeed spam" refers to the precision, while the recall would be "high share of spam emails is indeed classified as spam" and the false alarm rate would be "low share of no-spam emails is classified as spam / high share of no-spam emails is classified as no-spam"

d) **1 point** was awarded for correct axes labels.
4 points were awarded for correct points on the ROC curve (0.5 points per point, no point for (0,0) and (1,1), capped at 4 points). 1 point was deducted if the coordinates for an axis were missing (2 points deducted if both axis coordinates missing).
1 point was awarded for identifying the point corresponding to the cutoff-value correctly. **1.5 points** was awarded for interpreting it as a poor choice, since a modified cutoff-value could increase one rate <u>while not decreasing the other rate</u>. Partial credit was awarded for an incomplete interpretation.

e) **1 point** was awarded for identifying this statement as false. **1 point** was awarded for an explanation why the slope cannot equal the metric (comparing the formulas, counterexample, etc.). Partial credit was awarded for incomplete approaches or missing reference to the slope.

Problem 5 – Decision Trees

a) Common mistakes were not to use a binary split for the numeric attribute, rounding errors, and counting errors.

b) Gain ratio as stated in the exercises was slightly off. We apologize for this mistake. Therefore, we lowered the hurdle for reaching full credit for this exercise. Most essential was to state the gain ratio formula and to apply it to the new data of eleven instances.

- c) – e) are in random order. For the exercise of "You build a decision tree with pruning activated [….]", we were looking for alternatives to pre- or post-pruning: Thus, both didn't count as answers. Otherwise, grading should be self-explanatory.

Problem 6 – Causality

(No additional comments provided, grading scheme is explicitly provided in the exam questions.)