**Datacamp**

1. In the Datacamp exercise, Logistic Regression for Churn Prevention, it is mentioned that threshold can be empirically checked by using costs and the rewards of the campaigns. How can we define these rewards and costs?
2. (<https://campus.datacamp.com/courses/machine-learning-for-marketing-analytics-in-r/logistic-regression-for-churn-prevention?ex=9>) In this chepter, Pseudo R-squared statistics are mentioned. Is the interpretation with adjusted r-squared different? For example, it is taken as good if the value is more than 0.5. However adjusted r-squared of say 0.6 is not so good.
3. (<https://campus.datacamp.com/courses/supervised-learning-in-r-classification/chapter-3-logistic-regression?ex=5>) It is mentioned that although there can be different models with the same AUC, their shapes can differ. What is the difference in the blue and red AUC in 3:20. How can we make decisions to chose? Which one is best for which cases?
4. Datacamp- Dummy variables, missing data, and interactions. When one-hot encoding is applied to a categorical variable, do we need to add both variables to the regression model or do we just add one?
5. Datacamp video linked below (0:33). Can pseudo R^2 statistics (McFadden, Cox & Snell, Nagelkerke) give very different results? Would it be enough for only one to give a good result? <https://campus.datacamp.com/courses/machine-learning-for-marketing-analytics-in-r/logistic-regression-for-churn-prevention?ex=9>
6. Datacamp video linked below (4:20), there is a graph which shows the overfitting. I couldn’t understand this graph. Is this the sigmoid function? <https://campus.datacamp.com/courses/machine-learning-for-marketing-analytics-in-r/logistic-regression-for-churn-prevention?ex=9>

A picture containing man

Description automatically generated

1. Datacamp Course (Chapter - 3) GLM function (Logistic Reg.) exclude observations with missing values in training data. What if we have missing values in test data or in a real life case? Does it make any predictions for it or just exclude them again?
2. Let us assume that the pay off coefficients are concrete about return customer or not. Is it always best to select the minimum pay off's threshold value. Or should we take the trade off in accuracy into account? Reference: Datacamp Churn Analysis Part : Finding the Optimal Threshold
3. (Datacamp Application churn prevention). The answer of the question is that the main reason of using churn prevention is convincing existing customers to buy again and stay loyal to the online shop however, why it could not be used for the possible customers? We have the details of the customer insides so; it could be easily used for gaining more other customers.
4. In the DataCamp chapter, I did not understand the difference between the two models below. While regression works with all of the columns in logitModelFull, logitModelNew calculated by stepwise regression. when we look at the accuracies of these two models, they are approximately the same. Could the effect of the stepwise regression on accuracy be so insignificant?

In sample fit full model

A screenshot of a cell phone

Description automatically generated

In sample fit restricted model

A screenshot of a social media post

Description automatically generated

1. In DataCamp Course: “Supervised Learning in R: Classification”, Chapter 3 (Logistic Regression) Exercise 1; after fitting the model we are required to summarize the model object. In the summary, what does:
   * 1. Null deviance
     2. Residual Deviance
     3. AIC
     4. Number of Fisher Scoring iterations

tell us? What does AIC stand for? How do we interpret the model according to this summary? What inferences should we retrieve form the values of these fields?

1. On Chapter 3 (Logistic Regression) of Data Camp Assignment (Supervised Learning in R), we are asked a question on Exercise 7. Which ROC is better?

A close up of a map

Description automatically generatedSince three of the AUCs are close to each other, the correct answer is “more information is required”. I believe the correct answer should be the blue one with the highest AUC, at least theoretically. Can you please help me differentiate theory and practice?

1. On Chapter 2 (Churn Prevention in Online Marketing) of Data Camp Assignment (Machine Learning for Marketing Analytics in R) on 9th content (In sample model fit and thresholding, between 0:27 and 0:55) we are told that Pseudo R-squared values are interpreted as follows:

A picture containing table

Description automatically generated

In Statistics courses, we learn that R-squared values which are around 0.7 and 0.8 are accepted as good measures while we claim a value between 0.4 and 0.5 is good? What is the logic behind this?

1. Datacamp: Supervised Learning in R: Classification – Logistic Regression – Automatic Feature Selection (2:15) .The backward deletion and forward stepwise methods, which are automatic feature selection methods, can come up completely different solutions and features. In such a case, should we choose the way the model performs well at the end of the process, or the way it contains features that are better suited to our goals, even if the performance is relatively low? In addition, are these two methods much preferred?
2. Supervised Learning in R: Classification/ Chapter 3: Logistic Regression / Automatic Feature Selection /2:36. <https://campus.datacamp.com/courses/supervised-learning-in-r-classification/chapter-3-logistic-regression?ex=13>. It is said that “Statisticants also raise concerns about the fact that stepwise model violates some of the principles that allow a regression model to explain data as well as predict. Of course if you only care about the predictive power this may not be a very big concern” I could not clearly understand what is told in this paragraph. Would we discuss this in the class?
3. Supervised Learning in R: Classification/ Chapter 3: Logistic Regression/ Automatic Feature Selection /Building a Stepwise Regression Model. <https://campus.datacamp.com/courses/supervised-learning-in-r-classification/chapter-3-logistic-regression?ex=15>. The plot shows the x axis as Specificity, in other sources I see it is shown False Positive Rate as (1- Specificity) which one is true?

A close up of a map

Description automatically generated

1. At the same example how can we interpret what step\_model did the calculation? In the R console I see a part of the calculations as follows: A screenshot of a cell phone

   Description automatically generated
2. In our datacamp course, (chapter2 page 8) , narrator explains the meaning of hypothesis test for newsletter feature and adds because of it’s p value less than 0,05 we reject the null hypothesis and assume that newsletter does have significant effect on return of a customer. With doing this,are we assuming that all the features fit into a normal distribution or even it can’t have significant effect on return of a customer with different distribution type?

A screenshot of text

Description automatically generated

1. While iterating the regression model and changing the included features, is there any way to prioritize minimization of false positive or false negative values rather than prioritizing accuracy in stepwise regression.
2. As i understand, setting our threshold for training / test set under %50 probability(%25 lets say) to get more accurate true positives or eliminate false positives means that we are giving label 1 if an observation have %30 chance to be a label 1. Isn’t that a sign for overfitting?

A screenshot of a cell phone

Description automatically generated

1. ( https://campus.datacamp.com/courses/supervised-learning-in-r-classification/chapter-3-logistic-regression?ex=3 ) In this exercises, line 8 they defined treshold 0, 0.0504 How can we determine this value when we determine the threshold values in our models?
2. Datacamp: Logistic Regression for Churn Prediction. Finding the optimal threshold. Why are there using confMatrix[2,2] for true positive? İsn’t it [1,1]?? I didn’t understand that and couldn’t get any point from here. I also googled it.
3. Is backward stepwise method more reliable with many features than forward stepwise method? I think selecting first dominant feature from more than 30 features is less reliable. (Reference: Logistic regression datacamp video : Feature Selection part)
4. Machine Learning for Marketing Analytics in R/ Logistic Regression for Churn Prevention/In sample modelf’t and Thresholding 3:25 . <https://campus.datacamp.com/courses/machine-learning-for-marketing-analytics-in-r/logistic-regression-for-churn-prevention?ex=9>. The relation between payoffs and return of the customers are not clear. Truly predicting the churn brings 5 euro and falsely prediction of the churn brings 15 euro loss. Why the other cases does not effect the payoff?
5. Machine Learning for Marketing Analytics in R/ Logistic Regression for Churn Prevention /Finding the optimal Threshold example/ <https://campus.datacamp.com/courses/machine-learning-for-marketing-analytics-in-r/logistic-regression-for-churn-prevention?ex=12>. It says “You could see that the optimal threshold is 0.3” Why?

**Customer Analytics with R**

1. Customer Analytics with R – Reasons for Churn (Page 49).A customer who has never had any problems during his / her membership can say a low value in his / her answer to the NPS question as a result of a single event he / she may experience. In such cases, should the person's answer be accepted as it is, or can the effect of it be reduced by values such as the small number of complaints of the customer in question during his membership?
2. Customer Analytics with R – How to implement this algorithm using R (Page 51).There is a risk table added to the churn analysis code in the example. What is the role of this code in this analysis? What purpose is it used for?
3. ( In article customer-analytics-with-r page 49, title Reasons for churn) Is enough for detect reasons for churn NPS and weighted factor or are we need to do different analysis ?
4. On W1 material (Page 49) NPS is not enough so that we apply Churn. I did not get the idea behind this. I believe, net promoter score is a concept where it can be utilized in feature selection and could be a good interpreter for Churn detection or at least understanding consumer motivation. Which aspects of marketing analytics Churn prediction unlocks that NPS cannot depict?
5. In PDF document: “Customer Analytics with R”, page 48; it is noted that Churn Analysis aims to divide customers into 3 categories: active, inactive and "about to churn". I have two questions regarding this quote:
   1. All the other material regarding this week’s session, we are instructed to employ Logistic Regression for Churn Analysis, which is only good for when we have dichotomous target variables. However, according to the above quote, Churn Analysis may require 3 target variables. Are we sure that Logistic Regression is a good technique for Churn Analysis? Which other techniques can we use for Churn Analysis under these circumstances?
   2. The “inactive” and “about to churn” customers seems to display overlapping attributes, by definition. For example, users who are inactive for more than 6 months are considered to be churned in Telecom Industry. Should we disregard “inactive” category in such cases and have two categories instead? Or should we list the subscribers who are inactive for 6 months as “about to churn” and delist all subscribers who are inactive for more than 6 months?
6. In the book of “Customer Analytics with R”, what can be an example of natural churn rate which a company has unavoidably?

A screenshot of a cell phone

Description automatically generated

1. In Customer Analytics with R, page 48, it is mentioned that the churn analysis aims to identify active, inactive and “about to churn” customers. How inactive and “about to churn” customers can differ? How should a company approach to these different groups?

**SPSS Pdf**

1. (<https://www.dropbox.com/s/b3hn1st29a957bj/W7%20-%20spss10.pdf>) It is stated that logistic regression has a model just like linear regression in the begining of page 4; however, linear regression uses least-squared errors and logit regression uses maximum likelihood mothod. However, logit regression is not considered a linear model. Why is that so?
2. In pdf document on the dropbox link, it is mentioned that logistic regression calculates the probabilities of belonging one group rather than the other. It is also explained that the dependent variable should be dichotomous. I cannot imagine that the dependent variable is dichotomous but not categorical. Can we explain a little bit more?

A screenshot of a cell phone

Description automatically generated

1. In PDF document (Logistic Regression), page 4 and in YouTube video 4 (Statistics 101: Logistic Regression, Estimating the Probability); it’s explained that the membership to the target class is decided by the calculation of how much more likely it is that an observation could be a member of the target group rather than the other group. Doesn’t this suggest that observations with probability of membership around 0.50 (assuming the threshold is 0.50 – like 0.44 or 0.57) are assigned to one class or another with low accuracy? How can we feel comfortable about our model’s robustness in such cases? Isn’t this a red line for implementing Logistic Regression?
2. W7-spss10.pdf / Model Fit And The Likelihood Function. How is the log likelihood generated in a model? Is this connected to statistical significance?
3. In article W7 - spss10 page 2, title Assumptions Of Logistic Regression ) The article said us for logistic regression our groups need to be mutually exclusive and exhaustive. If we have small data groups are not mutually exclusive and exhaustive, should we eliminate them or we need to do apply another method.
4. My other question is from Spss.doc. What is the base of ln for Spss? Last week you said that the log base changes according to the number of items while processing trees. is the same thing true here?

**Youtube**

1. (Statistics 101: Logistic Regression Probability, Odds, and Odds Ratio video).Can the variables be discrete and continuous both in our regression model? For example, X1 is discrete; X2 is continuous. Could the model y use both X1 and X2? Should the continuous data be converted into discrete form?
2. (Statistics 101: Logistic Regression, Logit and Regression Equation).What if we have more than 2 dependent variables; could we use Barnaouli distribution?
3. In the video (9.24. minutes), lecturer said that if we use linear regression instead of logistic regression predicted values of the Dependent variables can be beyond 0 and 1. How can it be. We only have 2 variables that are binary.
4. Is there any relation between ROC and accuracy? (<https://www.youtube.com/watch?v=4Xw19NpQCGA)min=6:00>

**Others**

1. In the literature, I realised that mostly logistic regression used for predicting the customers churn. Why we use especially this algorithm? If we use ensemble methods like random forest or gradient boosted decision trees (XGBoost), don't we have more accuracy?
2. When we try to predict churn, are there any essential features which used in churn prediction or do these features change according to the project?
3. If we have more than 2 classes, can we still use AUC and ROC curve to analyse accuracy?
4. how can we choose the best regression model when we are looking adjusted R square? (<https://statisticsbyjim.com/regression/model-specification-variable-selection/>)
5. How change the train and test accuracy when we add new features into data?
6. Which feature selection method (forward or backward) is more suitable in real life applications? Is there any automatic way to select features like sklearn’s feature\_selection lib. ?
7. Since we use the cumulative response curve, not the ROC curve, to convey it to management, why don't we use the cumulative response curve in the same way when doing these experiments alone?
8. What is the best metric to evaluate a supervised binary classifier?
9. In the article confint() method uses %2.5 / % 97.5 partition? Why is that? Or can we change it to the %5 / % 95? We used 95 confidence interval in statistics class many times. Reference: <https://stats.idre.ucla.edu/r/dae/logit-regression/>
10. Why I get different results from SVM classification efficiency score ?