Notes from lectures

Data(“variable name”) = Data(“variable name”).replace({“1”:0, “0”:1})

Data.Churn = Data.Churn.astype(“category”).cat.reorder\_categoties([1,0]).cat.codes

1.Model fitting

- mathematical calculations behind

2.Model Evaluation

- model is already fitted and we have model parameters, hyperparameters and we want to see if it is good or not

- in classification problem evaluation metrics are accuracy, rec all, specificity

- accuracy does not care about distinction between 0 and 1, accuracy is high if at least one of them is high

3.Model Diagnostics

- assumptions

4.Model selection

- train/test split, crossvalidation

5.Model interpretation

- DT important features only as Dt is not interpretable

- LR is interpretable

ROC\_AUC

1) optimizes both recall and specificity

2) gives us treshhold independent predictions, as ROC\_AUC is calculated for all possible treshholds

Thus, optimization should be based on ROC\_AUC. But ROC\_AUC is not enough and when we have 2 models we have to calculate Accuracy, Recall and Spacificity.

For example when we have 2 models with very close ROC\_AUC, we also should use other metrics to decide which model to choose.

Interpretation

To interpret Logit we use marginal effects that shows us the effect on probabilities.