
Theoretical Exercises**Exercise 11.1: (Theoretical) Evaluation of models**

An insurance company uses a risk score for their new costumer. Depending on the score, the application for a contract is accepted or not. The data science team, created two models to predict the score.

On a test set, the results for both models and the desired values are given in the following table.

	1	2	3	4	5	6	7	8	9	10
true label	0.4	0.7	0.8	0.5	0.2	0.6	0.6	0.5	0.3	0.2
Model A	0.4	0.7	0.6	0.5	0.7	0.6	0.8	0.8	0.4	0.4
Model B	0.4	0.9	0.9	0.6	0.4	0.5	0.8	0.6	0.5	0.7

- a) Compute the mean squared error (MSE) and mean average error (MAE) for both models. Depending on these metrics, which model is the better one? Explain why!

For the insurance company, a risk score below 0.5 is acceptable (contract is accepted) and a risk score larger or equal 0.5 is not acceptable (contract is rejected).

- b) Transform the scores into a classification, where the True label results into the acceptance of the contract and False into the rejection.
- c) Compute accuracy, precision, recall and F1 score of the two models. Depending on these metrics, which model is the better one? Explain why!
- d) Can these models be combined to improve the results?

Practical Exercises

Exercise 11.2: (Practical) Ames Housing Classification

We consider the Ames Housing dataset and want to predict the SalePrice as classes (cheap, expensive) based on the given variables.

- a) Prepare the data for a machine learning model by
 - compute the new variables "years since built" and "years since remod/add"
 - categorize the values for SalePrice in cheap (0 - 160000) and expensive (160000 - ∞)
 - analyze all variables and add classes "other" for values with few occurrences
 - remove unneeded variables and variables with extreme imbalanced data (e.g. 2 classes, one with 99% of the data and the other with 1% of the data)
- b) Create a train / test splitting
- c) Create a decision tree classifier
- d) Analyze the quality of the classifier by computing the different measures given in lecture
- e) Train different versions of the classifier (by varying the parameters) and compare these

Hint: A decision tree classifier can be created with sklearn: <https://scikit-learn.org/1.5/modules/generated/sklearn.tree.DecisionTreeClassifier.html>