

SMA: Software Modeling and Analysis

Practical Session
Week 10

Assignment 10

Discussion

A10 - Exercise 01 | General knowledge

Exercise 1: General Knowledge (3.5 Points)

- What is the main purpose of RMSE?
 - It is a code inspection tool used for bug prediction.
 - It is a measure that represents the quality of the input data.
 - It is a measure to calculate an ML model's prediction error.
 - It is an efficient machine learning algorithm.
- Why are prediction error assessments important?
 - The prediction error helps choosing appropriate algorithms.
 - The prediction error helps optimizing configuration parameters.
 - The prediction error reveals potential memory bottlenecks during the analysis.
 - The prediction error supports users in choosing the correct colors for bars in bar charts.
 - The prediction error supports users in profiling the analysis.
- What is the f-measure exactly?
 - A measurement of a test's accuracy, *i.e.* considering only the precision, but not the recall.
 - A measurement of a test's accuracy, *i.e.* considering only the recall, but not the precision.
 - A measurement of a test's accuracy, *i.e.* considering the precision and recall.
 - A measurement of a test's precision, *i.e.* considering the precision and recall.
 - A measurement of a test's recall, *i.e.* considering the precision and recall.

- What is the difference between the terms *accuracy* and *precision*?
 - The term *accuracy* is well-defined, but the term *precision* is not.
 - The term *precision* is well-defined, but the term *accuracy* is not.
 - There is no difference between both terms.
 - They are used in different contexts, *i.e.* *accuracy* for systematic errors, and *precision* for statistical variability.
- Does the effort required to raise the test coverage increase linearly?
 - No, it rather shows behavior of a Dirichlet function.
 - No, it rather shows behavior of a signum function.
 - No, it rather shows behavior of an exponential function.
 - Yes, it shows linear behavior.
- What is true regarding precision and recall with respect to anti-virus software on a computer?
 - Confirmed and detected viruses are false negatives, false alarms are true negatives.
 - Confirmed and detected viruses are true negatives, false alarms are false positives.
 - Confirmed and detected viruses are true positives, false alarms are false negatives.
 - Confirmed and detected viruses are true positives, false alarms are false positives.

A10 - Exercise 02 | WEKA

Exercise 2: WEKA (2.0 Points)

- What are benefits of using the [WEKA](#) tool?
 - It is equipped with a simple to use interface and provides immediate feedback to the majority of your actions.
 - It supports only classification problems, but no regression problems.
 - It supports in the current version [capsule \(neural\) networks](#).
 - It has built-in data preprocessing facilities.
 - It can import various data formats for later processing.
 - The core functionality can be integrated into your own Java project.

A10 - Exercise 03 | Machine learning specifics

Exercise 3: Machine learning specifics (4.5 Points)

- Which of these statements are correct?
 - A binary classification problem solver assigns each input data entity one out of two possible output labels.
 - A classification algorithm can have real-valued (\mathbb{R}) or discrete input variables.
 - A multi-classification problem solver assigns each input data entity the same output label.
 - Classification is the task of predicting a discrete class label for each input data entity.
 - Classifying emails into the categories “spam” and “not spam” is a classification problem.
 - Regression is the task of predicting a real-valued output (\mathbb{R}) for each input data entity.
 - There is no preference of using a specific algorithm (class) for certain kinds of input data.
- What are the effects of an algorithm stuck in a local minimum?
 - In general, the results are not consistent even with the same input data.
 - The algorithm cannot stop until it finds a way out of the local minimum.
 - The effect of local minima must not be considered while choosing an ML algorithm (suite).
 - The results are legitimate and reproducible.

- Which statements are correct regarding underfitting and overfitting of a model?
 - Overfitting states a model that adapted too few peculiarities of the input data, hence the model did not gain (completely) its ability for specialization.
 - Overfitting states a model that adapted too many peculiarities of the input data, hence the model lost (partially) its ability for generalization.
 - Underfitting states a model that adapted too few peculiarities of the input data, hence the model did not gain (completely) its ability for specialization.
 - Underfitting states a model that adapted too many peculiarities of the input data, hence the model lost (partially) its ability for generalization.
- What were typical dataset partition sizes used for testing and training of the bug prediction models?
 - 1% test dataset, 99% training dataset
 - 30% test dataset, 70% training dataset
 - 70% test dataset, 30% training dataset
 - 99% test dataset, 1% training dataset

Assignment 11

Preview

A11 - Exercise 01 | Data in SDA

- What is a confusion matrix?
- Which criteria are important when extracting the data?
- Which are the three groups of software data?

A11 - Exercise 02 | Process of SDA

- What is a use case for software data analytics?
- What are types of software analytics problems?
- What is the recommended five-step guideline for working in software data analytics?

You have to attend the lecture to reveal such slides.*



**Disclaimer:*

The content that has been shown on this slide is irrelevant for the exam.