Beuth University of Applied Sciences
Media Informatics Online Master | Data Science | SS2019
Carolin Gellner

ML INTRODUCTION EXERCISE

EXERCISE 1: DOMINGO PAPER

WHAT IS THE MEANING OF *GENERALIZATION* IN THE PAPER?

The generalization is an aim of machine learning, which implies that a test dataset should be used to judge the success of a classifier, not just the training dataset. A test dataset is a dataset which the algorithm does not see during the traning.

WHAT ARE THE MANY FACES OF OVERFITTING?

Overfitting can have the following faces: low variance and high bias, low variance and low bias, high variance and high bias and high variance and high bias. Bias is the tendency of a learner to learn the same falsity again and again, and variance is the tendency to learn random things.

WHAT IS FEATURE ENGINEERING?

Feature Engineering is the design and specification of the right features for a class to achieve the best possible learning experience.

WHAT IS ENSEMBLE LEARNING?

Combining the strengthness of multiple weak classifiers to create a strong classififer.

WHY DOES MORE DATA BEATS CLEVER ALGORITHMS?

In essence all algorithms do the same. More data improve the ability to group the examples.

REFERENCES

[1] P. Domingos, „A Few Useful Things to Know about Machine Learning,“ 2012.