Discussion week 1 – 422

What is overfitting? Why is it a special problem when working with machine learning (data-adaptive) models? How can we protect against overfitting?

Overfitting is a model performance problem that overvalues the fit from training data which results in worthless predictions. This problem becomes apparent in machine learning as a reaction to data is a required step in the process. The provided data and learning algorithm can both be adjusted to protect against overfitting. Specifically, data can be cleaned, data sources can be expanded, and the number of parameters used for model calculation can be reduced (regularization) in order to minimize overfitting.s