

SOFTWARE

OVERFITTING AND MODEL DEGRADATION

31.08.2022

My beginner experience here isn't exhilarating - maybe others are suffering as well from poor models but never report it?

[During the training phase](#) the model tries to learn the patterns in data based on algorithms that deduce the probability of an event from the presence and absence of certain data. What if the model is learning from noisy, useless or wrong information? Test data may be too small, not representative and models too complex. As shown in the article linked above, increasing the depth of the classifier tree increases after a certain cut point only the training accuracy but not the test accuracy - [overfitting](#)! So this needs a lot of experience to avoid under- and overfitting.

What is [model degradation](#) or concept drift? It means that that the statistical property of the predicted variable changes over time in an unforeseen way. While the true world changes - maybe political or by climate or whatsoever - this influences also the data used for prediction making it less accurate. The computer model is static representing the time point when the algorithm has been developed. Empirical data are however dynamic. Model fit need to be reviewed in regular intervals and again this needs a lot of experience.

CC-BY-NC Science Surf , accessed 09.04.2026, [click to save as PDF](#)
