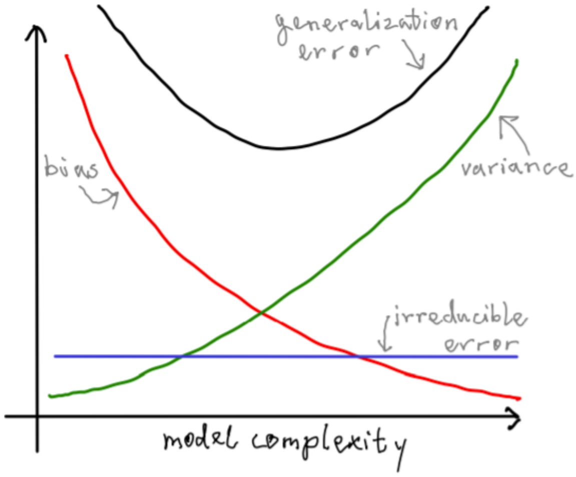

The graph is a visualization of Theorem 7.1 from the lecture notes on the bias-variance decomposition of the generalization error: GE = IE + $Bias^2$ + Variance. These are the four components:

- Irreducible Error: this is the error which is an intrinsic property from the data as noise and it cannot be changed. This is why the function on the graph is constant, as there is nothing that can be done to increase/reduce this error.
- Bias: deviation of the mean of the predictor with respect to the exact model. For a low model complexity, the bias is high, leading to underfitting of the data. In some sense the model is not "powerful" enough to find the patterns in the data well. Opposingly, for a bigger model complexity, the bias is high, leading to overfitting of data. This means that the model has adapted too much to the data, leading to higher error in data which it has not seen before.
- Variance: expected squared deviation around the mean (with respect to the training data) around the predictor. The variance works in contrast with the bias: less model complexity leads to small variance (underfitting), and bigger model complexity leads to more variance (overfitting). The variance shows how sensitive the model is to small fluctuations in the training data.
- Generalization error: This is the final measure of how accurately a model can predict data which it has not seen before. To achieve optimal results for the total GE, one must make a tradeoff between the bias and the variance with respect to the model, training data or similar factors. Factors that influence the model complexity are: whether it is linear or non-linear (non-linear is more complex), number of features in the training data (less features lead to smaller model complexity), number of data samples, specific parameters to a model (examples: *k* in KNN, number of layers in a neural network).