Subject: Re: Assessing A Machine Learning Method's Predictivity
Posted by thomas on Tue, 05 Apr 2022 12:56:04 GMT
View Forum Message <> Reply to Message

Hello Christophe,

I don't quite understand the question, but assume, there is a misinterpretation of the result. When you running the analysis, the data set is split into 10 fractions:

1: the oldest 10% of the data
2: the second oldest 10%
...
10: newest 10%

Then 9 models are generated using 10%, 20%, 30% ... 90% of the data (always the oldest)

Then every model is used to predict the Y value for the oldest fraction, which was not part of the model creation, e.g. for model 1 it is fraction 2, for model 5 it is fraction 6.

Then for all fractions with predicted data (2-10) a correlation graph is shown with predicted versus known Y-values. Graph 'fraction 8', for instance answers the question: If I had used built a model at the time, when I had 70% of the data and if I had used that model to predict Y-values for the next molecules to synthesize, how well would the prediction have been.

Does this explain it or did I misunderstand the question?

Thomas