

## 8.5: Can a Model Be Too Good?

Error sounds like a bad thing, and usually we will prefer a model that has lower error over one that has higher error. However, we mentioned above that there is a tension between the ability of a model to accurately fit the current dataset and its ability to generalize to new datasets, and it turns out that the model with the lowest error often is much worse at generalizing to new datasets!

To see this, let's once again generate some data so that we know the true relation between the variables. We will create two simulated datasets, which are generated in exactly the same way – they just have different random noise added to them.

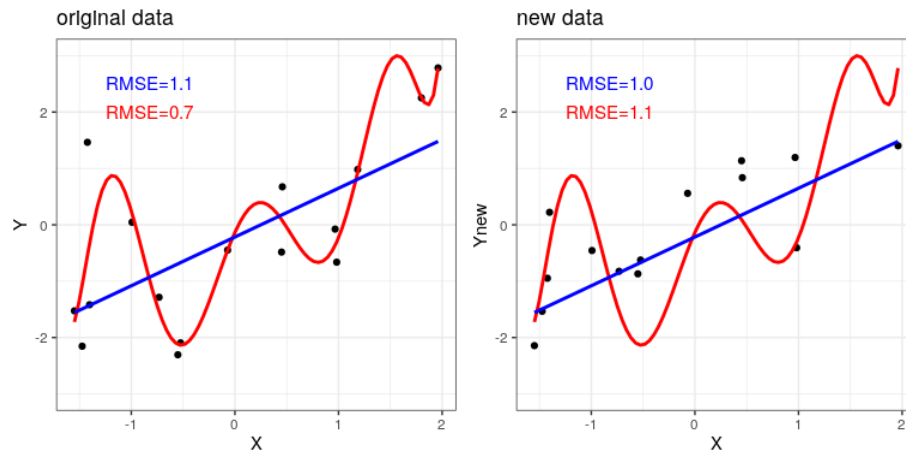


Figure 8.6: An example of overfitting. Both datasets were generated using the same model, with different random noise added to generate each set. The left panel shows the data used to fit the model, with a simple linear fit in blue and a complex (8th order polynomial) fit in red. The root mean square error (RMSE) values for each model are shown in the figure; in this case, the complex model has a lower RMSE than the simple model. The right panel shows the second dataset, with the same model overlaid on it and the RMSE values computed using the model obtained from the first dataset. Here we see that the simpler model actually fits the new dataset better than the more complex model, which was overfitted to the first dataset.

The left panel in Figure 8.6 shows that the more complex model (in red) fits the data better than the simpler model (in blue). However, we see the opposite when the same model is applied to a new dataset generated in the same way – here we see that the simpler model fits the new data better than the more complex model. Intuitively, we can see that the more complex model is influenced heavily by the specific data points in the first dataset; since the exact position of these data points was driven by random noise, this leads the more complex model to fit badly on the new dataset. This is a phenomenon that we call *overfitting*. For now it's important to keep in mind that our model fit needs to be good, but not too good. As Albert Einstein (1933) said: “It can scarcely be denied that the supreme goal of all theory is to make the irreducible basic elements as simple and as few as possible without having to surrender the adequate representation of a single datum of experience.” Which is often paraphrased as: “Everything should be as simple as it can be, but not simpler.”

This page titled [8.5: Can a Model Be Too Good?](#) is shared under a [CC BY-NC 4.0](#) license and was authored, remixed, and/or curated by [Russell A. Poldrack](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.