**Portfolio 2, Part 2**

**GitHub Link:** <https://github.com/nanmat/Assignment2-CogSci-Knowledge/blob/master/CogSci%20Knowledge%20part%202.Rmd>

*Nanna Kildahl Mathiasen, Alberte Baggesgaard Seeberg, Esther Dyngby Jensen and Simon Anneberg Merrild Hansen*

**Questions to be answered (but see guidance below):**

**1. Write a paragraph discussing how assessment of prediction performance is different in Bayesian vs. frequentist models**

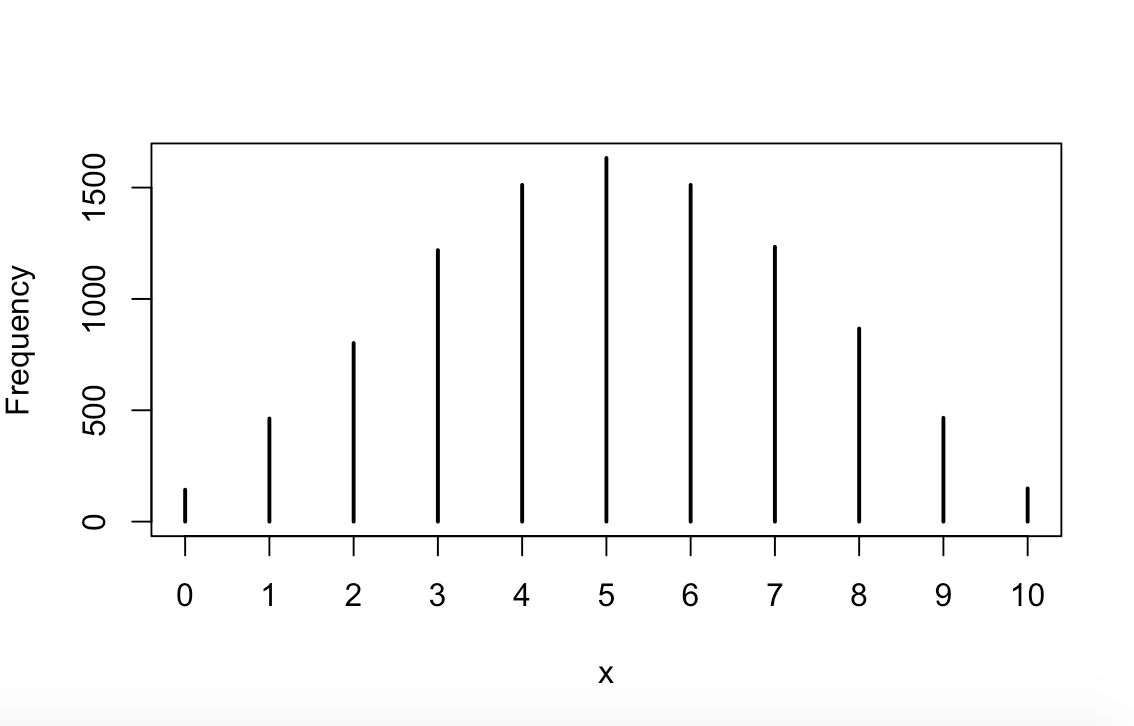
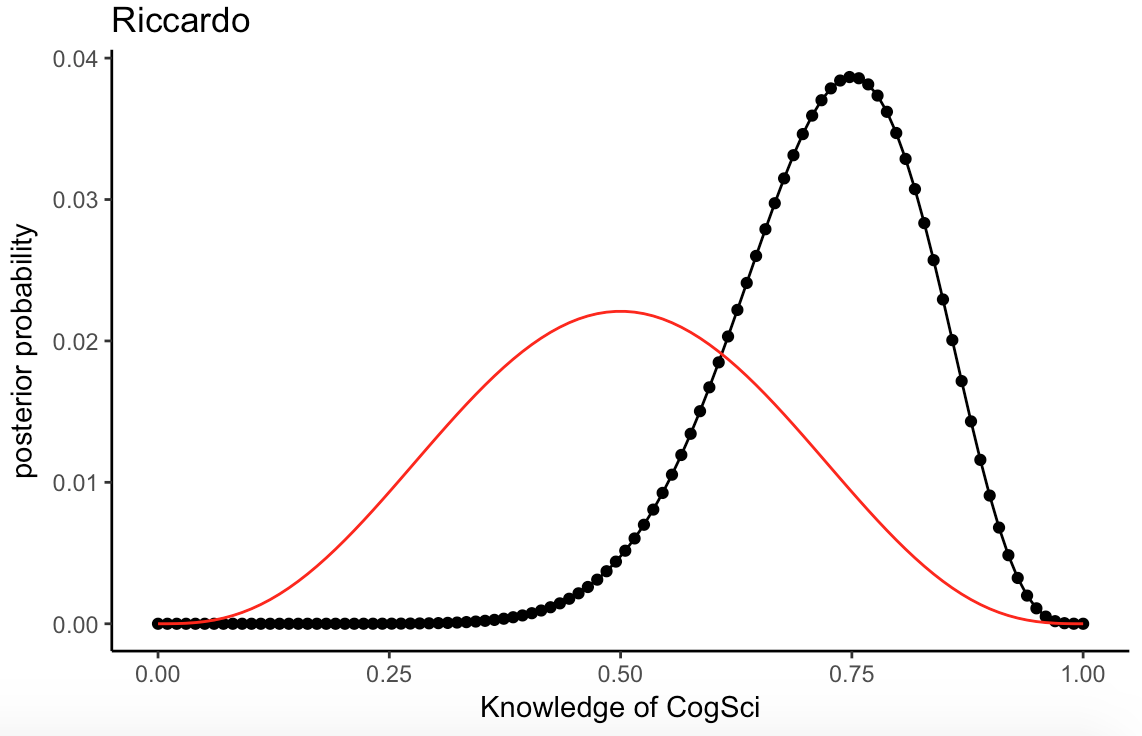
In both the Bayesian and the frequentist framework, we assess the error between the predictions of our model and the actual new data. The method of doing so is however quite different.

In the frequentist framework we generate predictions based on our model and compare these values to the empirical data, that the model is trying to describe. This difference can be expressed in e.g. RMSE (Root Mean Square of Error). The lower the RMSE value, the better the model performs.

In the Bayesian framework the posterior generated by previous data is used as a prior for the new data. If the prior resembles the posterior generated by the new data, we have a good model of our data. If the posterior is very different from the prior, then we have a bad model.

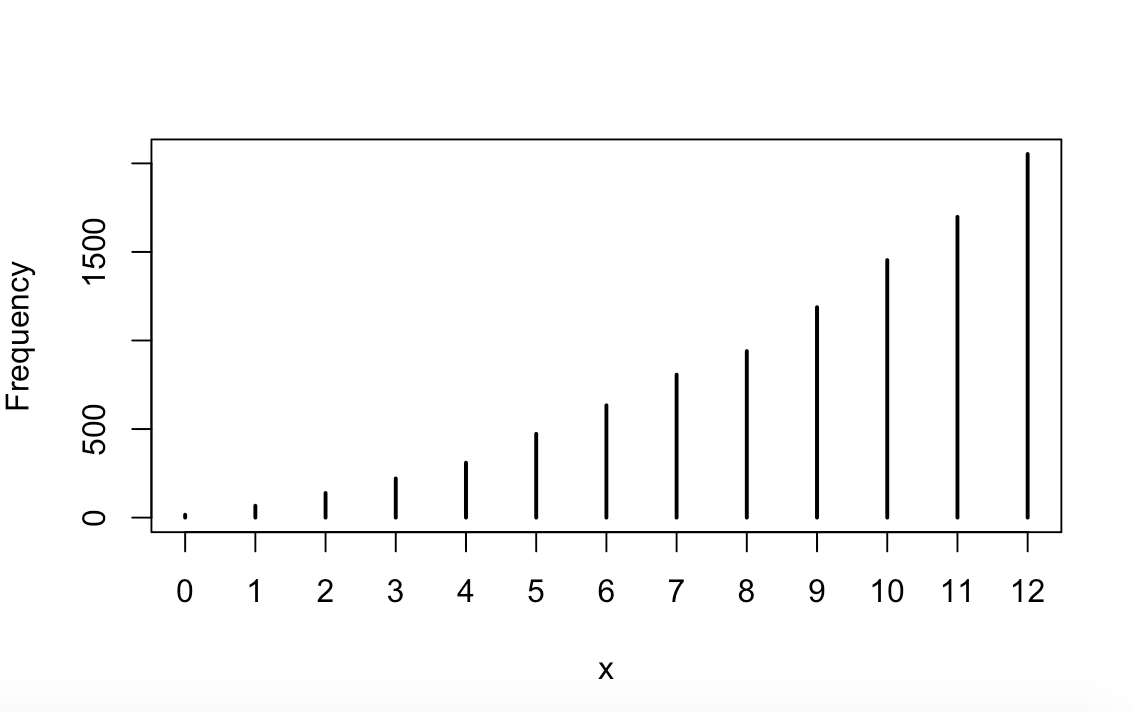
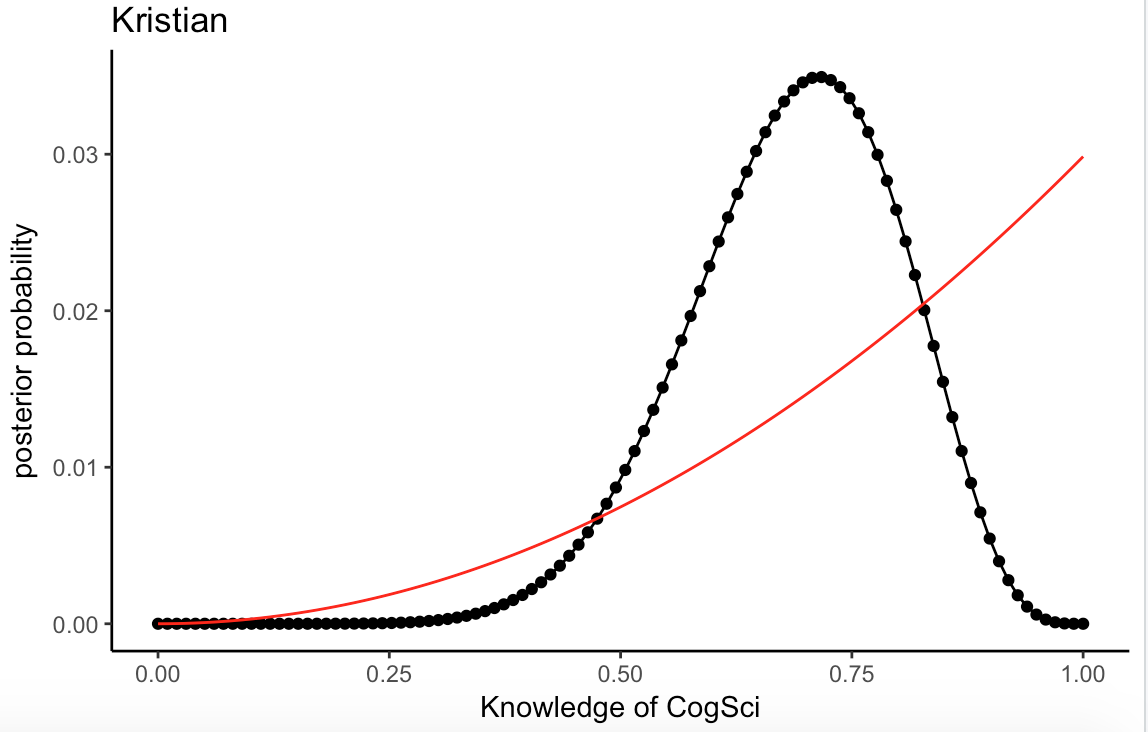
One of the major difference is that the values in the frequentist framework does not carry their uncertainty with them. This means that each value has the same weight. In the bayesian framework, we rely on distribution, which allow us to model the uncertainty in each value. If the prior is very uncertain (has a large spread), it is easily affected by new data, whereas if the prior is very narrow it requires more data to see a ‘significant’ change.

**2. Provide at least one plot and one written line discussing prediction errors for each of the teachers.**



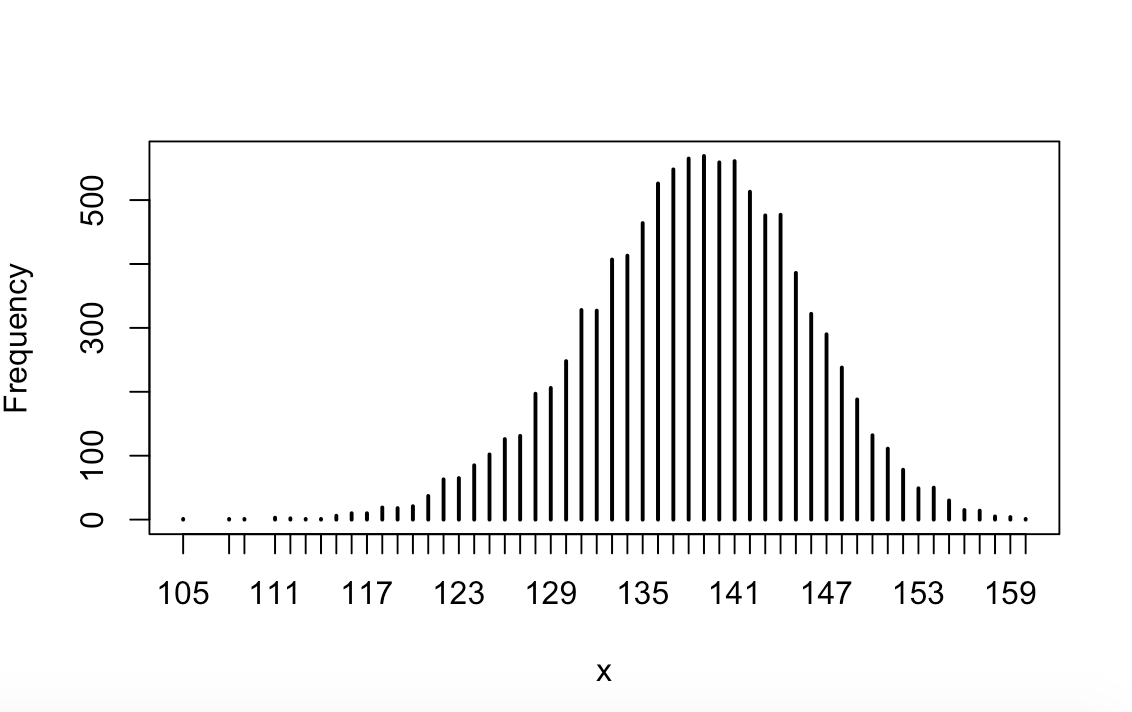
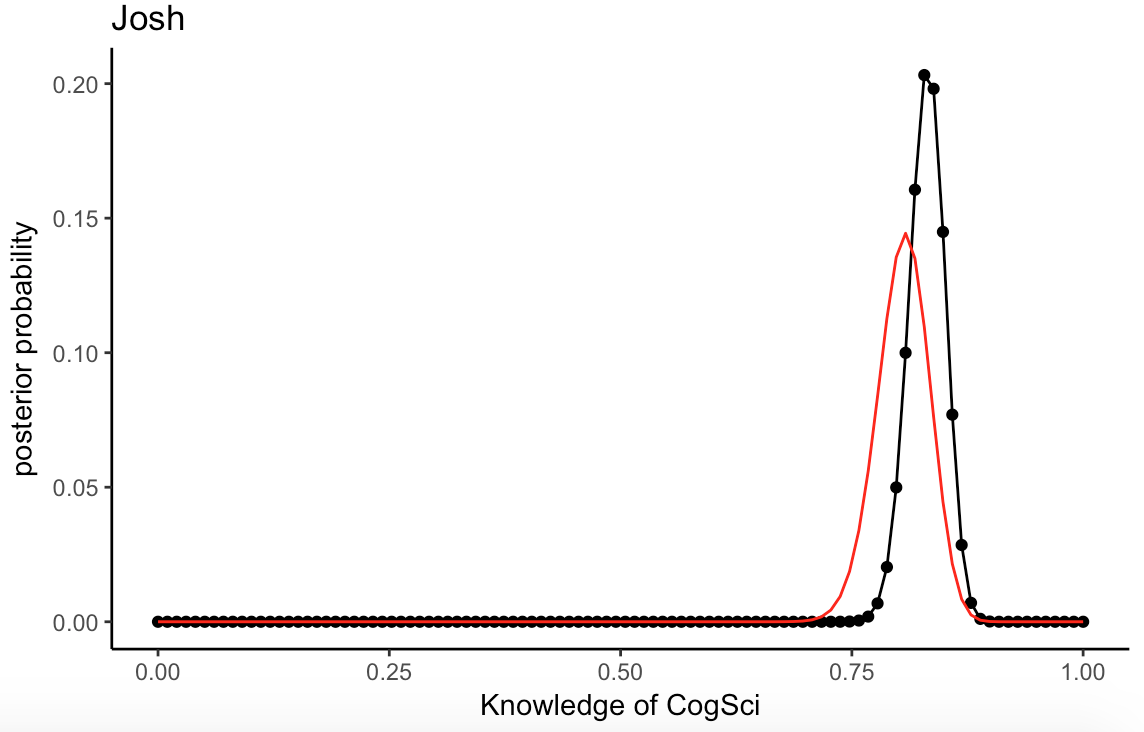
The data from last year is not good at predicting new data. Optimally the red (last year) and the black (this year) curve should be similar, but because of the few data points from last year, it leaves much uncertainty. The new data implies Riccardo to be better than chance, though this was expected to be most probable from the data from last year.

The histogram of the predictive posterior, showed in frequency of right answered questions, shows our model to not be a good predictor of the new data. Riccardo answered 9/10 correct in the new data but the histogram predicts that there will be more cases, where Riccardo answers 4-6/10 questions correct.  In addition, the PI (prob = 0.68) = [3, 7] shows that there is little chance that he would answer 9/10 correct given our model.



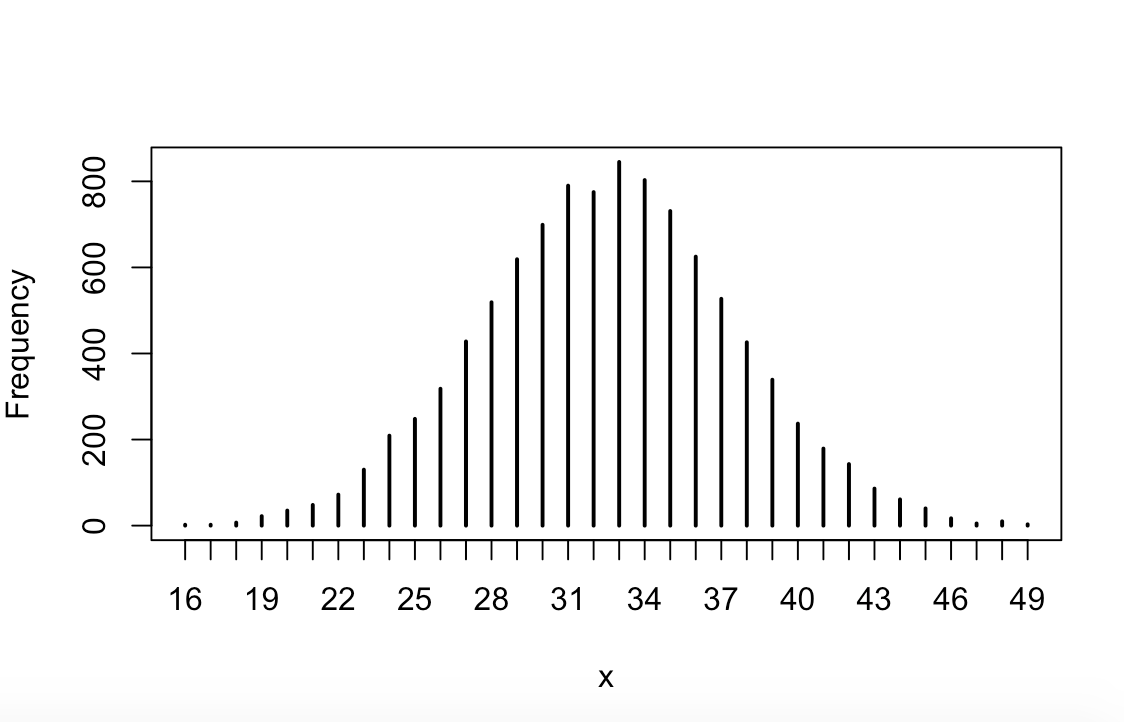
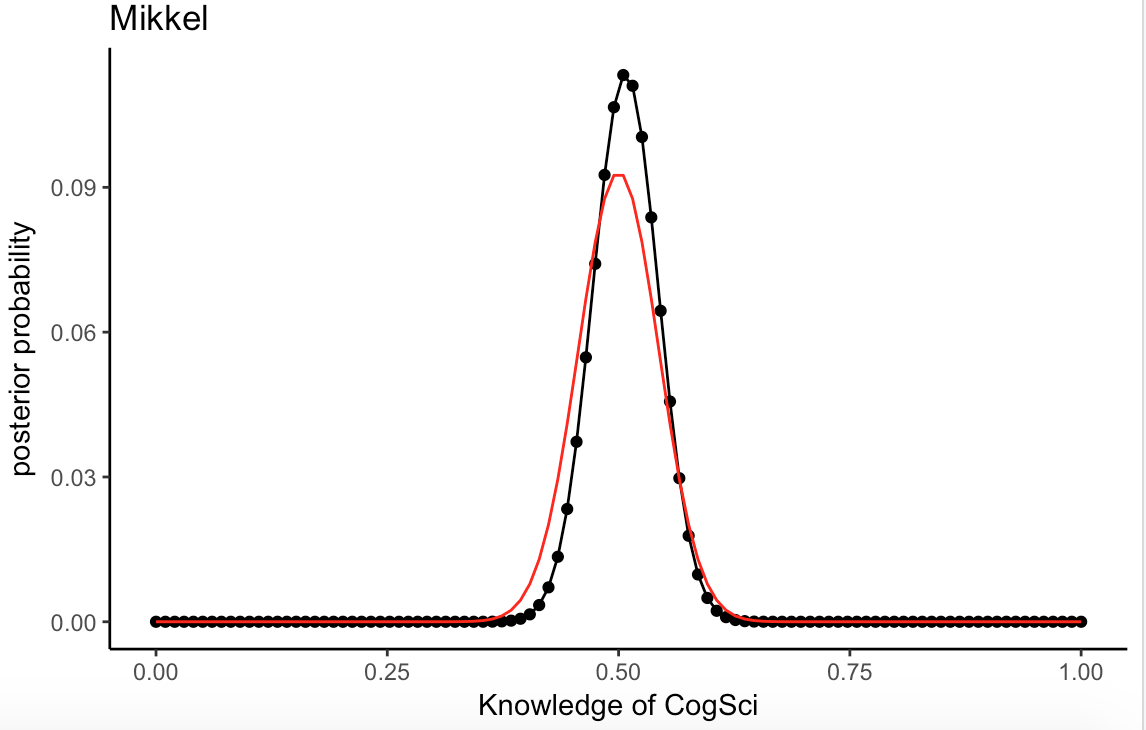
The data from last year is not good at predicting new data. The few data points from last year left much uncertainty. The new data gives a more certain estimate of Kristians actual knowledge of cog-sci.

The histogram of the predictive posterior shows our model to not be a good predictor of the new data. Kristian answered 8/12 correct in the new data but the histogram predicts that there will be more cases, where Kristian answers 11-12/12 questions correct, making him a cogsci superstar. <3 However, according to the PI (prob = 0.68) = [6, 12]  there is a good chance that he would answer 8/12 correct given our model, which shows that the model actually performs ok in predicting the data. This is probably mostly due to the high uncertainty of the prior given that he only answered two questions in the old data.



The data from last year is predicting the new data quite well. Josh answered many questions last year, which gave a more certain estimate. The new data lies close to what was expected. The new data has both increased the probability of Josh' knowledge of cog-sci to be well above chance, as well as increased percentage of actual knowledge of cog-sci.

The histogram of the predictive posterior, shows our model to be better at predicting the new data, although not very good. Josh answered 148/172 correct in the new data and the histogram predicts that there will be most cases, where Josh answers 135-145/172 questions correct. Looking at the PI, 148/172 is almost within the PI (prob = 0.68) = [131, 146], meaning that the model does an okay job at predicting the new data.



The data from last year is predicting the new data quite well. Like Joshua, Mikkel answered a lot of questions last year. The new data confirms the assumption, that Mikkel's knowledge of cog-sci is around chance level. The uncertainty has decreased a tiny bit.

The histogram of the predictive posterior shows our model is good at predicting the new data. Mikkel answered 34/65 correct in the new data and the histogram predicts that there will be most cases, where Mikkel answers 31-36/65 questions correct. In addition, 34/65 is within the PI (prob = 0.68) = [28, 37], meaning that the model does a good job at predicting the new data.