

MITOCW | MIT15_071S17_Session_2.4.05_300k

In this video we'll try to make predictions for the 2012-2013 season.

We'll need to load our test set because our training set only included data from 1980 up until the 2011-2012 season.

So let's call it `NBA_test`.

And we'll read it in the same way as always, `read.csv("NBA_test.csv")`.

All right, so now let's try to predict using our model that we made in the previous video, how many points we'll see in 2012-2013 season.

Let's call this `PointsPrediction`.

And so we use the `predict` command here.

And we give it the previous model that we made.

We'll give it `PointsReg4`, because that was the model we determined at the end to be the best one.

And the new data which is `NBA_test`.

OK, so now that we have our prediction, how good is it?

We can compute the out of sample r-squared.

This is a measurement of how well the model predicts on test data.

The r squared value we had before from our model, the 0.8991, you might remember, is the measure of an in-sample r-squared, which is how well the model fits the training data.

But to get a measure of the predictions goodness of fit, we need to calculate the out of sample r-squared.

So let's do that here.

We need to compute the sum of squared errors.

And so this here is just the sum of the predicted amount minus the actual amount of points squared and summed.

And we need the total sums of squares, which is just the sum of the average number of points minus the test actual number of points.

So the r-squared here then is calculated as usual, 1 minus the sum of squared errors divided by total sums of squares.

And we see that we have an r squared value of 0.8127.

We can also calculate the root mean square error the same way as before, root mean squared error is going to be the square root of the sum of squared errors divided by n, which is the number of rows in our test data set.

OK and the root mean squared error here is 196.37.

So it's a little bit higher than before.

But it's not too bad.

We're making an average error of about 196 points.

We'll stop here for now.

Good luck with the homework.