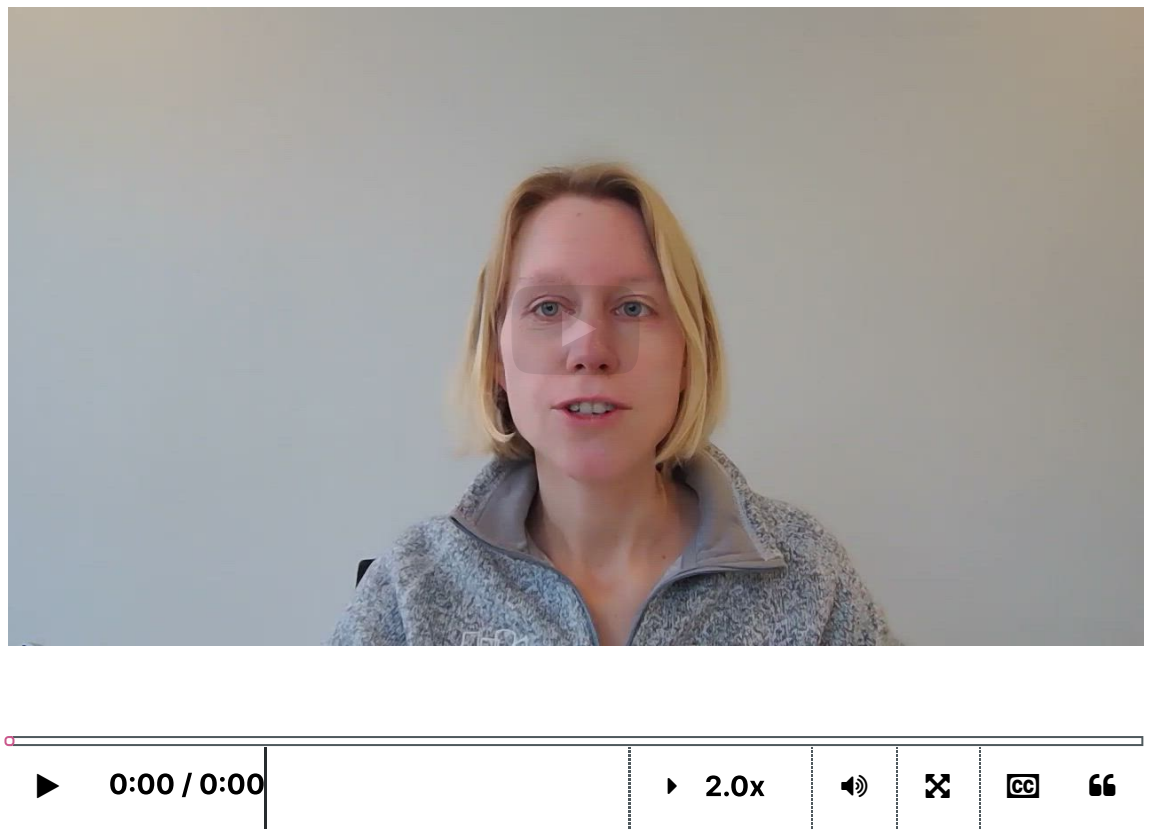
5. Information criterion

Bookmark this page

Exercises due Nov 10, 2021 17:29 IST Completed

Information criterion

Start of transcript. Skip to the end.



Prof Jegelka: We saw partial autocorrelation was

a very good way to actually estimate the model

order that an autoregressive model that should have,

so basically how far back in time should

Video
Download video file

Transcripts
Download SubRip (.srt) file
Download Text (.txt) file

We have learned so far how to use ACF and PACF to determine appropriate model structure to fit a time series. In this part, we will study two generic ways to determine which model is more appropriate among a few model candidates: **information criterion** and **cross validation** .

Information Criterion is a generic way to select among different model candidates. It provides a simple to calculate metric that considers how good does a model fit data (likelihood value) and the model complexity (often the number of the parameters in the model).

Akaike Information Criterion (AIC) is one of these criterions. AIC is calculated as:

$$AIC = 2k - 2ln(L)$$

where k is the number of parameters in the model, n is the number of the observations in the dataset and L is the likelihood value of a given data set.

Models with smaller AIC values are preferred to the models with larger AIC values. Smaller AIC values are associated with smaller number of model parameters (thus the model is less complex) and a better fit to the data (larger likelihood value).

AIC and BIC

4/5 points (graded)
Recall the definition of the Akaike Information Criterion (AIC) in the lecture:

$$AIC = 2k - 2ln(L)$$

Another commonly used information criterion is Bayesian Information Criterion (BIC). BIC can be calculated with the the following formula:

$$BIC = kln(n) - 2ln(L)$$

where k is the number of parameters in the model, n is the number of the observations in the dataset and L is the likelihood value of a given data set.

Consider two models of different complexity (A and B) to model a time series of length 1000 (n=1000). Model A has 10 parameters and the likelihood value is 10. Model B has 20 parameters and the likelihood value is 10^6.

What is the AIC value for model A? (Enter an answer correct to at least 2 decimal places for all questions below).

15.3948298140119 ✔ **Answer:** 15.39

What is the AIC value for model B?

12.3689788840715 ✔ **Answer:** 12.37

What is the BIC value for model A?

64.4723826038333 ✔ **Answer:** 64.47

What is the BIC value for model B?

110.524084463714 ✔ **Answer:** 110.52

Now think more generally about AIC and BIC, select all correct statements:

- [ ] AIC is a better criterion than BIC when n is small
- [ ] BIC is a better criterion than AIC when n is small
- [x] A complex model that is selected based on AIC may not be selected based on BIC ✔
- [ ] A simple model that is selected based on AIC may not be selected based on BIC ✔

✖

Solution:

Options A and B are not correct. No criterion is strictly better than the other criterion. It all depends on whether the assumptions made about these criterions are more appropriate under the current circumstance. Usually complex models are more preferable with AIC than with BIC, since BIC usually has a larger penalty on model complexity.

Submit You have used 3 of 3 attempts

Answers are displayed within the problem

Discussion

Topic: Module 4: Time Series:Introduction to Time Series Analysis 3 / 5. Information criterion

Hide Discussion

Add a Post

All Posts

Rolling Forecasting Origin

discussion posted 2 months ago by **ranitchatterjee0308**

Isn't the process tiring? I mean suppose we have T=100, and run it from T= 80, it'll take a lot of time to get the results, instead we could just split the entire stamp in to smaller sets, completely random, and split the data smaller pieces into train & test data, calculate the error for each set and evaluate the results. Won't it be a good idea?

This post is visible to everyone.

Add a Response

1 response

trungaero
2 months ago

I think the good thing is that you would have more chance to validate your model considering that your sample has limited number of data points. Furthermore, chopping off the data the way you describe then it could be the case that you can't use your data to fit a model with inherently high order.

Add a comment

Showing all responses

Add a response:

Preview

Submit

Previous Next

edX

About
Affiliates
edX for Business
Open edX
Careers
News

Legal