A question on cross-correlation & correlation coefficient [duplicate]

Asked 13 years, 6 months ago Modified 13 years, 6 months ago

Viewed 28k times

9

This question already has answers here:

Closed 13 years ago.

Possible Duplicate:
Matlab Cross correlation vs Correlation Coefficient question

When I cross correlate 2 data sets `a` and `b` (each 73 points long) in MATLAB and graph it, it appears like a triangle with 145 points. I'm confused between the correlation coefficient and the triangle-like graph when I plot the cross correlation output which ranges from +/- 1.

math **matlab** **data-analysis** **cross-correlation**

Share

Improve this question

Follow

edited May 23, 2017 at 12:09

Community Bot

1 • 1

asked Jun 8, 2011 at 20:13

user772225

2 Answers

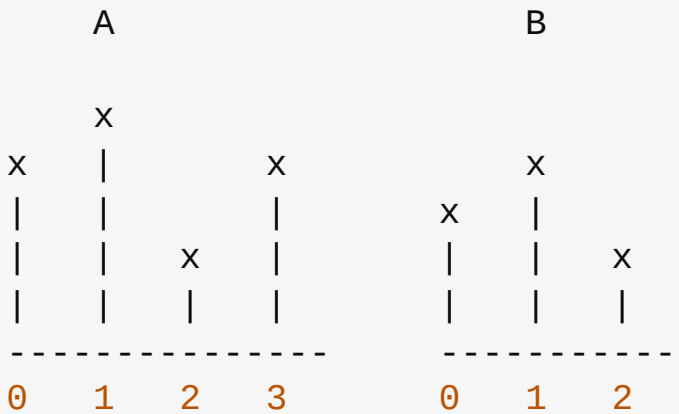
Sorted by: Highest score (default)

46

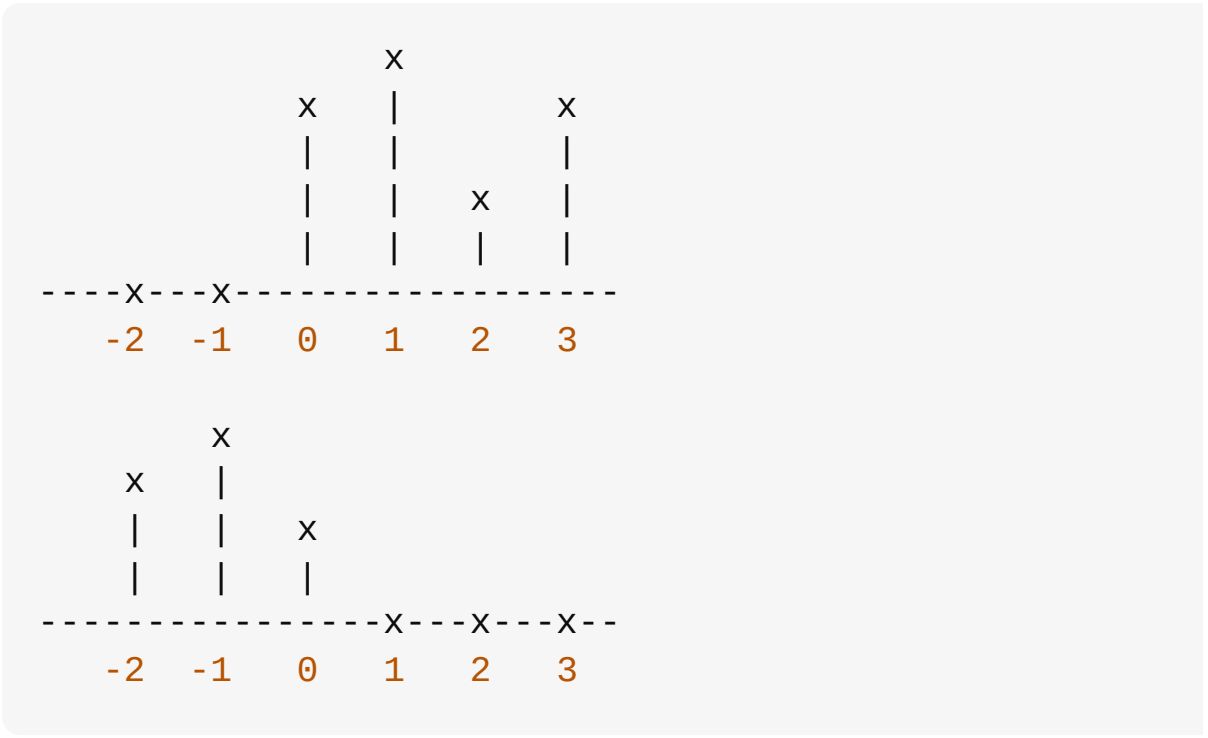
I seriously think you need to read up more on cross-correlation functions & correlation coefficient from a statistics book, because your confusion here is more fundamental than related to MATLAB. Unless you know what you're dealing with, you cannot make sense of what MATLAB gives you, even if you get the program right.

CROSS-CORRELATION:

Here is what you do in a cross correlation. Consider data `A` and `B` as follows

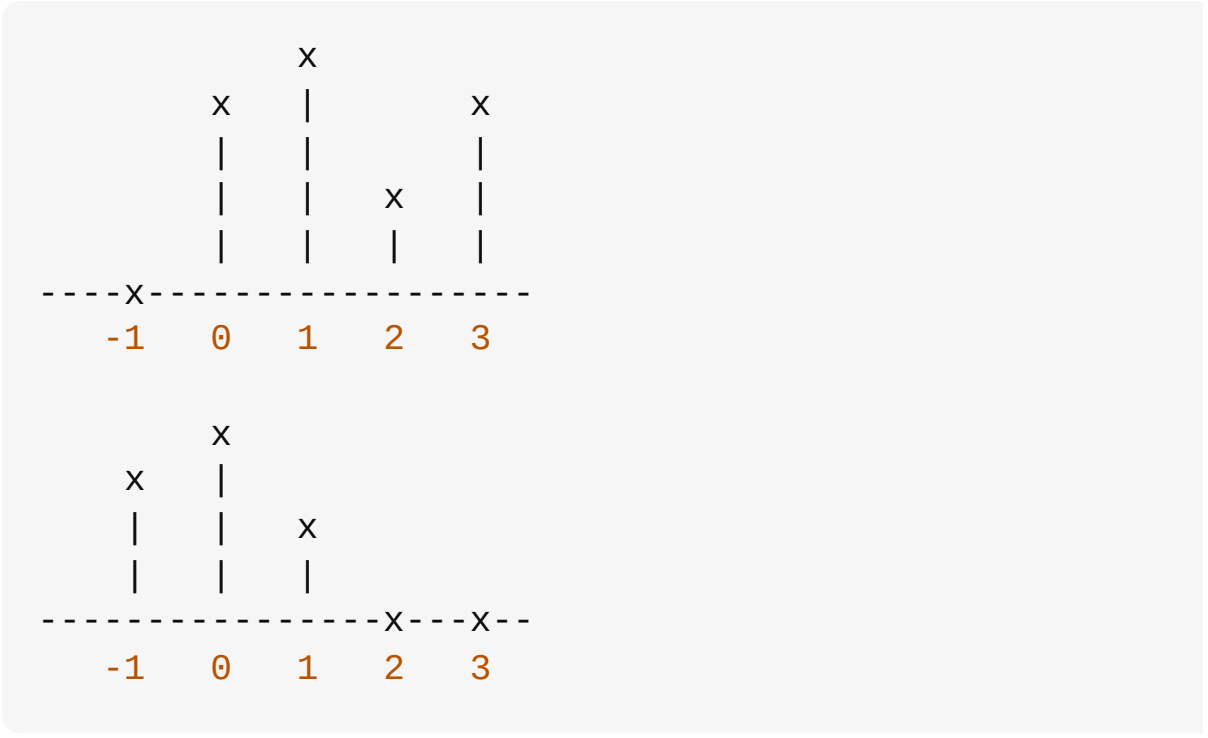


You then take `B` and slide it all the way to the end, so that the last point of `B` and the first point of `A` are aligned:



You fill in zeros where ever the data does not exist i.e., in this case, `B` beyond 0 and `A` before 0. Then you multiply them point wise and add, giving `0 + 0 + 3 + 0 + 0 + 0 = 3` as your first point in the cross-correlation.

Then you slide `B` one step to the right and repeat



giving `0 + 9 + 4 + 0 + 0 = 13` as the second point in the cross-correlation. You keep doing this till you slide `B` all the way to the other end of `A`.

The resulting vector is `length(A)+length(B)-1`, the -1 being because we started with an overlap at 0, so it's one point less. So here you should get `3 + 4 - 1=6` points in the cross-correlation and in your case, you should get `73 + 73 -1 = 145` points.

As you can see, the value of the cross-correlation vector at any point, need not be within ±1. The cross-correlation has a maximum when the two data vectors are "most alike". The "offset" of the peak from zero gives an indication of the "lag" between the two datasets.

CORRELATION COEFFICIENT

The correlation coefficient (I'm assuming Pearson's) is a mere number defined as

```
                Covariance(A,B)
r = ------------------------------
    ____________________________________
  \|Covariance(A,A)*Covariance(B,B)
```

where `Covariance(A,A)` is better known as `Variance(A)`. This is a quantity that can range from `-1` to `1` (as for why it has to be between ±1, look up Cauchy-Schwartz inequality)

NOTE:

While you can most certainly calculate the cross-correlation of two data vectors with unequal data points, you *cannot* compute their correlation coefficient. The notion of covariance is a measure of how two variables/datasets change *together* and is not defined for unequal datasets.

Share Improve this answer

Follow

edited Jun 20, 2020 at 9:12

Community Bot

1 • 1

answered Jun 8, 2011 at 21:39

abcd

42.2k • 7 • 84 • 99

1 So does it just show how similar 2 graphs are? By the way, thanks for the lengthy informative answer – user772225 Jun 8, 2011 at 21:47

2 @Tyler31: A better way to say it would be that it measures the similarity of two vectors/datasets/timeseries as a function of lag... – abcd Jun 8, 2011 at 21:51

1 @Tyler31: BTW, the cross-correlation in MATLAB returns values between -1 and 1 *only* if you ask it to do so with the `'coeff'` option. – abcd Jun 8, 2011 at 21:57

Thank you for the help with my confusion of terms – user772225 Jun 9, 2011 at 17:29

6 Thank you so much, Lorem, for your great explanation of how cross correlation works. I know this is an old thread (and a duplicate for that matter), but this answer is probably the best that I've seen. – rj2700 Jun 12, 2013 at 10:02

0

Have you read what that function returns?
http://www.mathworks.com/help/toolbox/signal/xcorr.html

> `c = xcorr(x,y)` returns the cross-correlation sequence in a **length** `2*N-1` vector, where `x` and `y` are length `N` vectors `(N>1)`.

`2*73-1=145` so that checks out. And the formula right below it explains why.

Share Improve this answer Follow

answered Jun 8, 2011 at 20:20