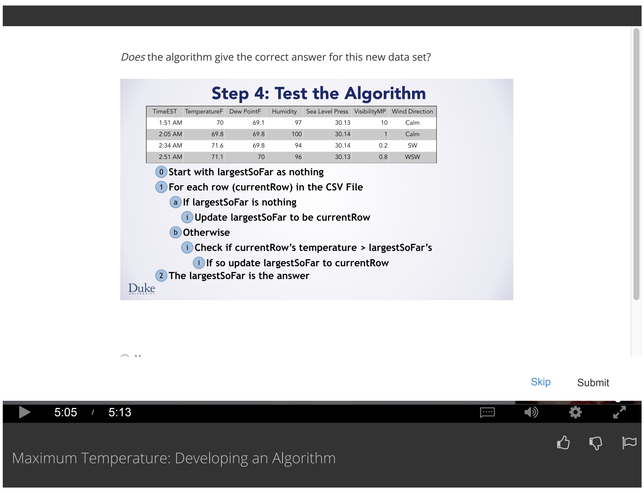

Have a question? Discuss this lecture in the week forums.

Interactive Transcript

Search Transcript

English

0:03
Welcome back. The first piece of solving the problem of finding the highest temperature in a year, which has its data spread across hundreds of files, is to find the maximum temperature in one day, which always has one file.

0:17
As always, we're going to start by working a small example by hand in a step by step fashion. Here we have six rows of data to work with.

0:28
The first thing you might do is look at the first row, and in particular at its temperature, which is 30 degrees Fahrenheit. That is the maximum we have seen so far, which you will wanna keep track of.

0:40
If you were just working this out with a small data set, you might just remember it in your head. But we will draw a red box around it to be explicit.

0:49
You might then go through the rest of the data in a similar fashion, looking at each row and deciding whether it is the largest you have seen so far or not.

0:58
After you looked through all the rows, you have found your answer. Which in this data file, is the last row.

1:04
Okay. Now that we have worked an instance of the problem by hand, it is time to write down exactly what we did in a step by step fashion. The first thing we did was look at the first row. In particular, its TemperatureF column. Next, we noted that it has the largest temperature so far.

1:21
The next step was to look at the second row. Its temperature is not greater than the largest temperature we have noted so far.

1:29
Then we looked at the third row.

1:31
Its temperature is 30.9, which is larger than the largest we have seen so far. So we updated our largest so far to be the third row.

1:41
For the fourth row, we did very similar steps. Saw that 32 was larger than the largest so far. And updated our largest so far.

1:52
The fifth row was not larger than the largest so far.

1:57
And the sixth row was larger than the largest so far. So we updated our note of what was the largest. That was the last row.

2:07
So we are ready to give the answer. In this case the sixth row was our answer.

2:13
Now that we have all of those steps written down for this particular instance of the problem we are ready to find patterns and generalize.

2:21
The first thing you might notice, is that you are doing similar, but not quite the same things, for each row of the CSV file. As you can probably guess by now, you will eventually write code, which loops over the rows to solve this problem. However, before you can do that you need to think about the differences, and find ways to make them the same.

2:39
The first difference you might notice is that for the first row, we just noted that it was the largest so far. But, for later rows, we compared the row to what we had previously noted down as our largest so far. The first row is a bit unusual here because we have nothing else to compare it to. We did something implicit that we did not write down. We check if our largest so far was nothing, or something first. We'll need to incorporate that into our generalized steps.

3:09
The other difference you might notice is that sometimes we updated what we recorded as the largest so far, while other times we did not. We marked the first rows update in purple as we just discussed how it is different from the others. And mark the steps in red where we did not update the largest so far, and in green where we did. What is the pattern?

3:30
It is when the current rows temperature is higher than the largest so far's temperature. Thinking through these patterns leads us to the following thoughts on how to decide when to update the largest so far.

3:43
If the largest so far is nothing, meaning we don't have one yet, then the current row is the largest so far. Otherwise, if the row's temperature is greater than the largest so far's temperature, then the current row is the largest so far.

3:59
After thinking through that, we can express our algorithms in terms of, for each row in the CSV file.

4:06
For each row, which we will call currentRow, you will want to decide how to update the largestSoFar variable, as we have just discussed.

4:16
We have not said anything about what largestSoFar starts as, so we should be sure to put that in here. We mentally glossed over this while we were writing down our steps, but we implicitly started with it as nothing before we began looking at each row. We should write that down in our algorithm here.

4:35
The last step, which we did write down, was to give our 6th row as our answer after we finished looking at each row.

4:44
Is the answer always gonna be the 6th row?

4:47
No.

4:49
It is always going to be the largestSoFar, the row that we have been keeping track of as we worked through the data.

4:57
We should test this out before we try to write our code. Try it out on these four rows of data. Does the algorithm give the right answer? Yes it does. We are now more confident that we wrote our algorithm correctly so we are ready to turn it into code.

Downloads

Lecture Video mp4
Subtitles (English) WebVTT
Transcript (English) txt
Lecture Slides pdf

Would you like to help us translate the transcript and subtitles into additional languages?