Dan Siegel

DS510

1. (3 points)

According to a 2017 Kaggle survey, 86.5% of data scientists said they used Python at work. By comparison, only 16.2% said they used Java and only 11.6% used C/C++.

Give three reasons why you think a data scientist would choose to use Python instead of Java or C/++.

1. Python is considered portable, and it is able to be used across multiple environments, from windows, to Linux, to OS X, to IBM mainframes.
2. Ease of Use: python is often referred to as pseudo code as it able to achieve the same result with less code. Specifically, one that I personally like is the need to not explicitly declare a type.
3. Ability to run without a separate compiler: python has a compiler built into the program as opposed to other languages such as C++. This allows for rapid prototyping and much shorter feedback loops.

2. (3 points)

Give three reasons a data scientist might choose to use a language like C or Java instead of a language like Python or R.

1. Java, C, C# all execute the code a lot a lot faster.
2. The ability to have strict typing. As much as it is an asset in python, it is a detriment for those who prefer it and in situations that it is needed.
3. The code could be better supported for that platform. For example, I know a company that uses data science to fight fraud, and they specifically use C#. They chose that because they are on a windows environment, and the code is better supported in Visual Studio.

3. (3 points)

What is the difference between a CPU and a GPU?

What advantages does a CPU have over a GPU? What advantages does a GPU have over a CPU?

* A CPU is a general purpose processing unit used to perform tasks on your computer. A GPU is another processing unit specifically optimized to produce images. The main difference is the number of cores and the ability for parallel processing. A CPU can perform a single calculation much faster than a GPU, a GPU can perform that same calculation in parallel much faster than a CPU.

4. (3 points)

Most programming languages, including Python, Java, C, and R, are Turing-Complete.

Using your own words, explain in a few sentences what it means for a programming language to be Turing-complete.

* A Turing Machine essentially boils down to a device which can complete a list of instructions to give you a result. Complex problems are broken down into simple instructions, and no matter how complicated the problem is, and given a long enough instruction set, the Turing Machine should be able to calculate the solution. A programming language is considered Turing-Complete when it can do anything that a Turing Machine can do.