Nimish Mishra — 39/CSE/18051/362

Fundamentals

① **Data** — Data is generally something that describes the entity about which it is collected in the first place. Concept of data is so abstract that is it is not informative to the user. Only when data is analyzed, the analysts learn something about the entity for which data has been collected, and it becomes information about the entity. Further principles based on the information yield to knowledge about the entity. Data (with respect to data science) is usually textual (structured) or in form of images, sound (unstructured) etc. For example; numbers about age and height of people is data. The information that height increases with age and then stops increasing after a certain point can be gained by analysing those numbers. And knowledge would be linking this information to information about growth hormones to better understand Nature.

② **Science** — is the systematic process of building information from observations around. Due to its broad nature, it is usually difficult to gain an understanding of the definition. It is, hence, useful to look at a couple of examples. For instance, natural sciences seek to gain knowledge about Nature in general. This process involves making observations and thinking of hypothesis that might generalize those observations. A systematic approach of developing experiments to confirm those hypothesis is applied, and theories are developed. This

is the essence of natural sciences.

③ Data Science — Merger of data and science.
Data science involves acquiring knowledge about an entity by collecting and analysing data about that entity in a scientific manner. Data science is usually considered a specific subset of science that deals with hypothesis from data. For instance, we might want to look at patterns in stock market prices. We collect numbers about past prices (this is the unorganized, unanalyzed data) and develop some hypothesis that might explain its variation over time (this is science). We might come to accept or reject hypothesis based on ~~experiem~~ experiments (this is science), and if found acceptable, these hypothesis help predict future prices (something scientific hypothesis/theories do). So data science crudely means scientific process of extracting information from unanalyzed data.

④ Computers — Comparing vast majority of electronic devices an individual is surrounded with, there are some which can be broadly ~~sep~~ segregated from others because they are able to process information given to them and take actions. Formally, these are able to take some input, process it as defined through programming, and produce some output. They might even have some memory to prevent repeated processing of inputs. In this way, smart-phones, calculators, smart-watches, modern washing machines all have some portion of them working as a computer (taking in user input and directing the hardware based on some programmed computation performed). And the desktop workstation is certainly a misnomer, casually called 'computer'; it is just a 'type' of computer, not 'the' computer