

CSE 415

Assignment 1

Part B

Jerry Hong

yh47@uw.edu

1. My major is mathematics. And my major field is data science. *Data science is a "concept to unify [statistics](#), [data analysis](#), [machine learning](#) and their related methods" in order to "understand and analyze actual phenomena" with data. It employs techniques and theories drawn from many fields within the context of [mathematics](#), [statistics](#), [computer science](#), and [information science](#). Turing award winner [Jim Gray](#) imagined data science as a "fourth paradigm" of science ([empirical](#), [theoretical](#), [computational](#) and now data-driven) and asserted that "everything about science is changing because of the impact of information technology" and the [data deluge](#). In 2015, the [American Statistical Association](#) identified [database management](#), [statistics](#) and [machine learning](#), and [distributed and parallel systems](#) as the three emerging foundational professional communities. — —Data science, wikipedia.*
2. The computer does a good job at the statistics part. It has faster calculation speed than human, so that it can make the chaotic and dirty data clean. And its visualization allows it to display the data model directly. Its excellent calculation ability also allows it to do prediction of the data model.
3. What data scientists do is extracting [knowledge](#) and insights from structured and [unstructured data](#). — —Data science, wikipedia. The job of human is mainly

translating data to meaningful words and sentences — — conclusions. Finding what data, and the distribution and the pattern of data, to summarize the rules and laws.

4. i. A real data scientist. ii. Concise introduction about data science and the ability to communicate with people. iii. The knowledge storage of statistics.
5. The ability to analyze different types of data, such as distributions.
6. AI can not only analyze the tendency and distribution, but what is behind the data also.
7. Explain the Normal Distribution to normal people, which is the fundamental knowledge of statistics, also data science.