Jessica Gardner

CS 1030

Unit 5 Lesson 1 Data Tools:

This unit is mainly about the format of data in files and the uses of files. It talks about CSV files, which are represented differently than text files, spreadsheets, databases, or text files. The data can be used to find patterns in data. Different types of formatting of data helps patterns become more evident. One the data is formatted, it can then be useful. The organized data can be used to find averages, trends, medians, correlations, and other statistical manipulations.

Unit 5 Lesson 2 Big Data:

This unit is about big data. Big data can come from a variety of places. It comes from scientific research, consumer data, medical records or almost any other number of sources. The manipulation of big data can present different challenges. Giant data sets need efficient methods of storage as well as ways to help data scientists see patterns emerge. The collection of data also presents many ethical questions. How long should personal data be stored? Should an individual have access to the data that is collected about them? These are just some of the questions that arise. Parallel data processing, as well as distributed processing, are two popular methods to deal with large amounts of data.

Unit 5 Lesson 3 Bias in Machine Learning:

This unit focused on machine learning. Machine learning is when a computer program “teaches” itself whatever the designer of the program wants it to learn. There are multiple different ways that the machines can learn. Among those ways are reinforcement learning, unsupervised learning, and supervised learning. The design of the machines themselves are pivotal. Machines can be made out of neural networks. A network must be trained. The training requires massive amounts of data. Even in training a machine recognize rabbits in a picture requires an impressive amount of effort. The machine must be trained to recognize rabbits from every different angle and in every different type of lighting with every texture and color of background. The data must be labelled correctly. An incorrectly labelled datum can create problems as the machine learns from the datum without knowing the mistake. Machine learning can also be used as a predictive algorithm. These predictive algorithms can be used to predict the likelihood of repeat offenses of people convicted of previous crimes, decisions of companies to hire an individual or not. The ethics of the use of these predictions are hotly debated. Machine learning can also be used for facial recognition. Again, the need for huge amounts of correct data comes into play. The data given to the machine must include all different types and shades of faces as well as be able to distinguish thee face from its background.

Unit 5 Test Experience:

I enjoyed taking the unit test, as it was a good way to exercise my critical thinking skills. The test helped me realize I knew more about the topics of this unit than I realized. I got 8/9 questions correct, and the one that was incorrect was simply because I read the question incorrectly. I feel that I learned more from the unit test and exercises in the modules than the chapters themselves. Having to interact with the material in a meaningful way helped me more than simply reading.