Self-Review Questions

1.

* Data Science
  + Data Science is a field that is based around the extraction of knowledge from data, structured or not, using various scientific methods
* Data Mining
  + Data mining is the active process of extracting the useful data and knowledge from the collection of data
* Machine Learning
  + Machine Learning is the use of data and algorithms to “learn” how to imitate humans and surpass in speed and reliability

2.

a) Data Query

b) Data Mining

c) Data Query

d) Data Query

e) Data Mining

3.

Fully automated data mining is not desirable for the main reason human intervention will catch thing that the tool wont

4.

Data Mining would help uncover things that aren’t obvious and will make connections between things that could greatly determine outcomes

5.

One big reason for concern with data mining is privacy and with the amount of data collected on everyone it becomes a powerful tool that could end up in bad hands if not secured, with all the data on people things can be worked out about people that the said person does not want out in the world.

6.

a) Classification

b) Data Description

c) Estimation

Laboratory Questions

1.

Data mining is used is a variety of fields one such fields is fraud detection to help prevent fraudulent documents and various other documents

2.

* Data set: Forest Fires
* Awareness for forest fire protection
* The location of the individual
* Classification
* Likelihood of Forest fire in the area