AI Course Module 1 Notes

**1.2**

The overall goal of **Artificial Intelligence** is to get machines to behave intelligently. This can come in many forms, such as predictive algorithms and imitation of human behavior.

In some cases, the amazing feats accomplished by artificial intelligence can be categorized as displays of raw processing power rather than shows of true intelligence.

**Machine Learning** can be said to be the pursuit of true intelligence for machines. Though intelligence is a goal, true intelligence is encouraged by allowing the machine to learn from interactions with its environment. The more exposure a machine has to data, the better it will be at handling it.

Machine learning algorithms are oftentimes the fundamental building blocks of AI projects.

**Data exhaust** is the trail of data you leave behind daily, such as GPS locations any buying items. Data exhaust is often used heavily in machine learning.

**1.3**

**Data** is raw statistics and facts – they’re uncontextualized and, oftentimes, not particularly useful.

**Information** is data put into context. It meaningfully contextualizes the observed statistics, facts, or trends in data.

When categorizing whether something is information or data, you have to take into account what you’re being given and the context in which it’s being presented in.

**1.4**

Machine learning algorithms can be split into two categories – supervised and unsupervised.

**Supervised Learning** occurs when there is information available, but not enough information to learn until more data is gathered. Supervised learning has preexisting results that are labelled (training data sets) that can feed the algorithm data that will then be tested by test data sets. This data falls into two categories – dependent and independent.

**Independent data** is used as a target or a predictor of dependent data.

What’s yielded by **dependent data** relies on the results of the independent data.

Supervised learning trains the algorithm via training data, then tests its accuracy with test data. This process improves its ability to accurately respond to future data, called **generalization**

If test data isn’t flexible enough, the test data will **overfit** the algorithm, making it less effective practically.

**Unsupervised learning** occurs when the structure of the data set is unknown – in other words, unclassified and unlabeled.

Most data will have a mixture of labeled and unlabeled data within its data set.

**Classification** is when independent data is defined as a class label and has a definite discrete value. As a result, predictions made from such data are definite and discrete.

**Regression** is when ranges of data are stored using real numbers. Thus, predictions made from such data could also be in a range. In many cases, regression data can be changed to classified data via categorization.

**Structured data** has a high degree of organization, meaning each item falls into a particular type. This is typically displayed in a table format.

**Unstructured data** doesn’t conform to the rules of structured data. Though unstructured data may have some structure, some degree of nonconformity will be present.