COMP4431- Data Mining

Assignment 1

Megan Hoeksema

Write a paragraph or two defining each of these topics:

1. **Vectors**

In linear algebra, a vector is essentially a one-dimensional list of number. Vectors can also be thought of as columns of a matrix or a point in a multidimensional space. A vector can be denoted by a column of number enclosed in brackets. Vectors can undergo different operations such as addition, subtraction, and multiplication.

1. **Matrix**

A matrix is much like a vector, however, instead of being a one-dimensional list of numbers, a matrix is a multi-dimensional collection/table of numbers. Much like vectors, matrices can also undergo different operations such as addition and multiplication. Each column of a matrix is comprised of a vector, each sharing the same length. A few different special matrices include inverse matrices and identity matrices. The product of a matrix and its inverse is simply the identity matrix where a matrix of n x n dimension has a 1 across the diagonal for every element in the diagonal and a 0 elsewhere.

1. **Linear equations**

A linear equation is an equation in the form of ax + b = 0, where x is the variable and a, b are coefficients. Linear equations can also contain two variables, in the form of ax + by = c, where x and y are the variables and a, b and c are the coefficients. These coefficients are often in the set of real numbers. When performing operations on linear equations, everything that you perform on one side must be performed on the opposite side as well. This ensures the equation is balanced and maintains equality.

1. **Vector space in multiple dimensions**

A vector space is a non-empty set of vectors on which are defined two operations called addition and scalar multiplication. A vector space can be defined in multiple dimensions if at least one of the vectors cannot be written as a scalar multiplication of another vector in the set. Vector spaces have a set of rules that they must follow which include closure under addition, commutativity of addition, associativity of addition, additive identity, additive inverse, closure under scalar multiplication, distribution of scalar multiplication, distribution of summation of scalars multiplication, associativity of scalar multiplication, and scalar multiplication identity.

1. **Linear Transformations**

A linear transformation is written as T: V 🡪 W, where V and W are vector spaces. V is the domain of T, and W is the codomain of T. The following must hold true for any linear transformation: T(v1 + v2) = T(v1) + T(v2) and T(cu) = cT(u) for any scalar c. Linear transformations can be classified as surjective (onto) or injective (one-to-one). The transformation is said to be surjective if its range is equal to the codomain. It can be classified as injective is each element from the domain maps to a distinct element in the range and no elements in the range can be mapped to multiple elements in the domain.

A transformation can also be both surjective and injective, classified as bijective.

1. **Linear independence**

A set of vectors can be classified as linearly dependent when a nontrivial linear combination of the vectors is equal to the zero vector. Thus, if there exists at least one non-zero scalar, the vectors are said to be linearly independent. In order to establish linear independence of a matrix, the determinant must first be solved for. Then, the matrix is reduced to echelon form, until it can no longer be reduced.

1. **Eigenvalues and Eigenvectors**

The eigenvector of an n x n matrix A is a non-zero vector **v** such that Av = λv. The scalar, λ, is the eigenvalue, of which there is a non-zero solution. To determine if a given λ is an eigenvalue the equation (A – λI)v = 0 must hold true, where I is the identity matrix. Eigenvectors can be determined from the null space of the matrix. The null space of a matrix can be found by reducing the matrix to echelon form.

1. **Orthonormal bases and compliments**

Orthogonal denotes the perpendicularity of vectors. An orthonormal basis is an orthonormal set that is linearly independent. To determine the orthonormal basis, the vector must be normalized. The orthonormal complement is the set of vectors perpendicular to an original subspace.

1. **Compare and contrast data mining, machine learning and deep learning**

Data Mining is the process of analyzing large amounts of data with the goal of identifying patterns. Data mining is a very important process in business intelligence in that the data can be mined to find interesting relationships that can make important business decisions. Data mining can also be very important in industries that are client-facing such as marketing, sales, finance and HR. The core elements of data mining include machine learning and statistics. Data mining can be broken down into four core processes, those being, gathering, preparing, mining, and analysis/interpretation.

Machine Learning can be thought of as an application of artificial intelligence (AI). It focuses on developing programs that can access data and essentially learn from it. The model is developed based on a training set, then tested using a validation set. This is not something that is required in data mining. The main goal of machine learning is to automate processes with the idea of improving decision making. Some of the most common machine learning algorithms are online chatbots, self-driving cars and speech recognition.

Deep Learning is a subfield of machine learning in which the algorithms are inspired by the workings of the human brain. Deep learning tends to drive AI applications and services. The goal of deep learning is to improve automation to the point of not needing human intervention. Many products already incorporate deep learning, such as digital assistants and credit card fraud detection. Deep learning differs from machine learning in that machine learning structured, labeled data to make predictions, whereas deep learning eliminates these pre-processing.