Machine Learning
Lesson 1

Basics of Linear Algebra (LA) for Machine Learning (ML)

Lecturer: Dr. George Musumba
Email: george.musumba@dkut.ac.ke

Introduction

- Linear algebra is a pillar of machine learning.
- Linear algebra is a large field of study that has tendrils into engineering, physics and quantum physics.
- Only a specific subset of linear algebra is required, though you can always go deeper once you have the basics.

Assumptions from Logic Programming course

- You know your way around basic Python for programming.
- You may know some basic NumPy for array manipulation.
- Linear Algebra is necessary to deepen the understanding and application of machine learning.

Expected outcomes of studying LA

- What linear algebra is and why it is relevant and important to machine learning.
- How to create, index, and generally manipulate data in NumPy arrays.
- What a vector is and how to perform vector arithmetic and calculate vector norms.
- What a matrix is and how to perform matrix arithmetic, including matrix multiplication.
- A suite of types of matrices, their properties, and advanced operations involving matrices.

Expected outcomes of studying LA

- What a tensor is and how to perform basic tensor arithmetic.
- Matrix factorization methods, including the eigendecomposition and singular-value decomposition.
- How to calculate and interpret basic statistics using the tools of linear algebra.
- How to implement methods using the tools of linear algebra such as principal component analysis and linear least squares regression.

Outline of the Topic

- Foundation.
- NumPy.
- Matrices.
- Factorization.
- Statistics.

Foundation

- Linear algebra is a field of mathematics that is universally agreed to be a prerequisite to a deeper understanding of machine learning.
- Linear algebra is a branch of mathematics, but the truth of it is that linear algebra is the mathematics of data.
- Matrices and vectors are the language of data.
- Linear algebra is about linear combinations.
- A linear equation is just a series of terms and mathematical operations where some terms are unknown; for example:

$$y = 4 x + 1$$

Numerical Linear Algebra

- The application of linear algebra in computers is often called numerical linear algebra.
- “Numerical" linear algebra is really applied linear algebra.

Linear Algebra and Statistics

- Linear algebra is a valuable tool in other branches of mathematics, especially statistics.
- Usually students studying statistics are expected to have seen at least one semester of linear algebra (or applied algebra) at the undergraduate level.

Linear Algebra and Statistics

- Some clear fingerprints of linear algebra on statistics and statistical methods include:

➢ Use of vector and matrix notation, especially with multivariate statistics.

➢ Solutions to least squares and weighted least squares, such as for linear regression.

➢ Estimates of mean and variance of data matrices.

➢ The covariance matrix that plays a key role in multinomial Gaussian distributions.

➢ Principal component analysis for data reduction that draws many of these elements together.

Applications of Linear Algebra

- In his classical book on the topic titled “Introduction to Linear Algebra”, Gilbert Strang states the following applications of linear algebra
  - ➢Matrices in Engineering, such as a line of springs.
  - ➢Graphs and Networks, such as analyzing networks.
  - ➢Markov Matrices, Population, and Economics, such as population growth.
  - ➢Linear Programming, the simplex optimization method.
  - ➢Fourier Series: Linear Algebra for functions, used widely in signal processing.
  - ➢Linear Algebra for statistics and probability, such as least squares for regression.
  - ➢Computer Graphics, such as the various translation, rescaling and rotation of images.

Applications of Linear Algebra

- Another interesting application of linear algebra is that it is the type of mathematics used by Albert Einstein in parts of his theory of relativity. Specifically tensors and tensor calculus.
- He also introduced a new type of linear algebra notation to physics called Einstein notation, or the Einstein summation convention.

Linear Algebra and Machine Learning

- Linear algebra is a field of mathematics that could be called the mathematics of data.
- It is undeniably a pillar of the field of machine learning, and many recommend it as a prerequisite subject to study prior to getting started in machine learning.
- Linear algebra is the mathematics of data and the notation allows you to describe operations on data precisely with specific operators.

Examples of Linear Algebra in Machine Learning

- **Examples include:**
- ➢1. Dataset and Data Files
- ➢2. Images and Photographs
- ➢3. One Hot Encoding
- ➢4. Linear Regression
- ➢5. Regularization
- ➢6. Principal Component Analysis
- ➢7. Singular-Value Decomposition
- ➢8. Latent Semantic Analysis
- ➢9. Recommender Systems
- ➢10. Deep Learning

Dataset and Data Files

- In machine learning, you fit a model on a dataset. This is the table like set of numbers where each row represents an observation and each column represents a feature of the observation.
- For example, below is a snippet of the Iris flowers dataset (http://archive.ics.uci.edu/ml/datasets/Iris)

```
5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.0,1.4,0.2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
4.6,3.1,1.5,0.2,Iris-setosa
5.0,3.6,1.4,0.2,Iris-setosa
...
```

Dataset and Data Files

- This data is in fact a matrix, a key data structure in linear algebra.
- Further, when you split the data into inputs and outputs to fit a supervised machine learning model, such as the measurements and the flower species, you have a matrix (X) and a vector (y).
- The vector is another key data structure in linear algebra.
- Each row has the same length, i.e. the same number of columns, therefore we can say that the data is vectorized where rows can be provided to a model one at a time or in batch and the model can be pre-configured to expect rows of a fixed width.

Images and Photographs

- Perhaps you are more used to working with images or photographs in computer vision applications.
- Each image that you work with is itself a table structure with a width and height and one pixel value in each cell for black and white images or 3 pixel values in each cell for a color image.
- A photo is yet another example of a matrix from linear algebra.
- Operations on the image, such as cropping, scaling, shearing and so on are all described using the notation and operations of linear algebra.

One Hot Encoding

- Sometimes you work with categorical data in machine learning.
- Perhaps the class labels for classification problems, or perhaps categorical input variables.
- It is common to encode categorical variables to make them easier to work with and learn by some techniques.
- A popular encoding for categorical variables is the one hot encoding.
- A one hot encoding is where a table is created to represent the variable with one column for each category and a row for each example in the dataset.
- A check or one-value is added in the column for the categorical value for a given row, and a zero-value is added to all other columns.
- For example, the variable color variable with the 3 rows:

One Hot Encoding

-
```
red
green
blue
...
```
Might be encoded as
```
red, green, blue
  1,    0,     0
  0,    1,     0
  0,    0,     1
...
```

- Each row is encoded as a binary vector, a vector with zero or one values and this is an example of a sparse representation, a whole sub-field of linear algebra.

Linear Regression

- Linear regression is an old method from statistics for describing the relationships between variables.
- It is often used in machine learning for predicting numerical values in simpler regression problems.
- There are many ways to describe and solve the linear regression problem, i.e. finding a set of coefficients that when multiplied by each of the input variables and added together results in the best prediction of the output variable.

Linear Regression

- If you have used a machine learning tool or library, the most common way of solving linear regression is via a least squares optimization that is solved using matrix factorization methods from linear regression, such as an LU decomposition or an singular-value decomposition or SVD.
- Even the common way of summarizing the linear regression equation uses linear algebra notation:

$$y = A \times b$$

- Where y is the output variable A is the dataset and b are the model coefficients.

Regularization

- In applied machine learning, we often seek the simplest possible models that achieve the best skill on our problem.
- Simpler models are often better at generalizing from specific examples to unseen data.
- In many methods that involve coefficients, such as regression methods and artificial neural networks, simpler models are often characterized by models that have smaller coefficient values.

Regularization

- A technique that is often used to encourage a model to minimize the size of coefficients while it is being fit on data is called regularization. Common implementations include the L^2 and L^1 forms of regularization.
- Both of these forms of regularization are in fact a measure of the magnitude or length of the coefficients as a vector and are methods lifted directly from linear algebra called the vector norm.

Principal Component Analysis (PCA)

- Often a dataset has many columns, perhaps tens, hundreds, thousands or more.
- Modeling data with many features is challenging, and models built from data that include irrelevant features are often less skillful than models trained from the most relevant data.
- It is hard to know which features of the data are relevant and which are not.

Principal Component Analysis

- Methods for automatically reducing the number of columns of a dataset are called dimensionality reduction, with the most popular method being the principal component analysis (PCA).
- This method is used in machine learning to create projections of high-dimensional data for both visualization and for training models.

Principal Component Analysis

- The core of the PCA method is a matrix factorization method from linear algebra.
- The eigendecomposition can be used and more robust implementations may use the singular-value decomposition (SVD).

Singular-Value Decomposition

- Another popular dimensionality reduction method is the singular-value decomposition method or SVD for short.
- As mentioned and as the name of the method suggests, it is a matrix factorization method from the field of linear algebra.
- It has wide use in linear algebra and can be used directly in applications such as feature selection, visualization, noise reduction and more.

Latent Semantic Analysis

- In the sub-field of machine learning called natural language processing, it is common to represent documents as large matrices of word occurrences.
- For example, the columns of the matrix may be the known words in the vocabulary and rows may be sentences, paragraphs, pages or documents of text with cells in the matrix marked as the count or frequency of the number of times the word occurred.

Latent Semantic Analysis

- This is a sparse matrix representation of the text.
- Matrix factorization methods such as the singular-value decomposition can be applied to this sparse matrix which has the effect of distilling the representation down to its most relevant essence.
- Documents processed in this way are much easier to compare, query and use as the basis for a supervised machine learning model.
- This form of data preparation is called Latent Semantic Analysis (LSA) and is also known by the name Latent Semantic Indexing (LSI).

Recommender Systems

- Predictive modeling problems that involve the recommendation of products are called recommender systems, a sub-field of machine learning.
- Examples include the recommendation of books based on previous purchases and purchases by customers like you on Amazon, and the recommendation of movies and TV shows to watch based on your viewing history and viewing history of subscribers like you on Netflix.

Recommender Systems

- The development of recommender systems is primarily concerned with linear algebra methods.
- A simple example is in the calculation of the similarity between sparse customer behavior vectors using distance measures such as Euclidean distance or dot products.
- Matrix factorization methods like the singular-value decomposition are used widely in recommender systems to distill item and user data to their essence for querying and searching and comparison.

Deep Learning

- Artificial neural networks are nonlinear machine learning algorithms that are inspired by elements of the information processing in the brain and have proven effective at a range of problems not least predictive modeling.
- Deep learning is the recent resurged use of artificial neural networks with newer methods and faster hardware that allow for the development and training of larger and deeper (more layers) networks on very large datasets.

Deep Learning

- Deep learning methods are routinely achieving state-of-the-art results on a range of challenging problems such as machine translation, photo captioning, speech recognition and much more.
- At their core, the execution of neural networks involves linear algebra data structures multiplied and added together.

Deep Learning

- Scaled up to multiple dimensions, deep learning methods work with vectors, matrices and even tensors of inputs and coefficients, where a tensor is a matrix with more than two dimensions.
- Linear algebra is central to the description of deep learning methods via matrix notation.
- It is also central to the implementation of deep learning methods such as in Google's TensorFlow Python library that has the word “tensor" in its name.

NumPY-Arrays

- Arrays are the main data structure used in machine learning.
- In Python, arrays from the NumPy library, called N-dimensional arrays or the ndarray, are used as the primary data structure for representing data.
- We use N-dimensional array in NumPy for representing numerical and manipulating data in Python.

NumPy N-dimensional Array

- NumPy is a Python library that can be used for scientific and numerical applications and is the tool to use for linear algebra operations.
- The main data structure in NumPy is the ndarray, which is a shorthand name for N-dimensional array.
- When working with NumPy, data in an ndarray is simply referred to as an array.
- It is a fixed-sized array in memory that contains data of the same type, such as integers or floating point values.

NumPy N-dimensional Array

- The data type supported by an array can be accessed via the **dtype** attribute on the array.
- The dimensions of an array can be accessed via the **shape** attribute that returns a tuple describing the length of each dimension.

NumPy N-dimensional Array

- Array() function is used to create array
- The example below creates a Python list of 3 floating point values, then creates an ndarray from the list and access the arrays' shape and data type.

```
# create array
from numpy import array
# create array
l = [1.0, 2.0, 3.0]
a = array(l)
# display array
print(a)
# display array shape
print(a.shape)
# display array data type
print(a.dtype)
```

Running the example prints the contents of the ndarray, the shape, which is a one-dimensional array with 3 elements, and the data type, which is a 64-bit floating point.

```
[ 1.  2.  3.]
(3,)
float64
```

Functions to Create Arrays

- **Empty**
- The empty() function will create a new array of the specified shape. The argument to the function is an array or tuple that species the length of each dimension of the array to create.
- The values or content of the created array will be random and will need to be assigned before use. The example below creates an empty 3x3 two-dimensional array.
- Your specified array contents will vary.

```
# create empty array
from numpy import empty
a = empty([3,3])
print(a)
```

Functions to Create Arrays

- **Zeros:**
- The zeros() function will create a new array of the specified size with the contents filled with zero values.
- The argument to the function is an array or tuple that species the length of each dimension of the array to create.
- The example below creates a 3x5 zero two-dimensional array.

```
# create zero array
from numpy import zeros
a = zeros([3,5])
print(a)
```

Functions to Create Arrays

- **Ones**
- The ones() function will create a new array of the specified size with the contents filled with one values.
- The argument to the function is an array or tuple that species the length of each dimension of the array to create.
- The example below creates a 5-element one-dimensional array.

```
# create one array
from numpy import ones
a = ones([5])
print(a)
```

Combining Arrays

- **Vertical Stack**
- Given two or more existing arrays, you can stack them vertically using the vstack() function.
- For example, given two one-dimensional arrays, you can create a new two-dimensional array with two rows by vertically stacking them.
- This is demonstrated in the example below.

```
# create array with vstack
from numpy import array
from numpy import vstack
# create first array
a1 = array([1,2,3])
print(a1)
# create second array
a2 = array([4,5,6])
print(a2)
# vertical stack
a3 = vstack((a1, a2))
print(a3)
print(a3.shape)
```

Combining Arrays

- Horizontal Stack
- Given two or more existing arrays, you can stack them horizontally using the hstack() function.
- For example, given two one-dimensional arrays,

  you can create a new one-dimensional array or
- one row with the columns of the first and
- second arrays concatenated. This is demonstrated
- in the example below.

```
# create array with hstack
from numpy import array
from numpy import hstack
# create first array
a1 = array([1,2,3])
print(a1)
# create second array
a2 = array([4,5,6])
print(a2)
# create horizontal stack
a3 = hstack((a1, a2))
print(a3)
print(a3.shape)
```

Exercises

- Experiment with the different ways of creating arrays to your own sizes or with new data.
- Locate and develop an example for 3 additional NumPy functions for creating arrays.
- Locate and develop an example for 3 additional NumPy functions for combining arrays.

Further Reading

- Check the following URLs-
- https://machinelearningmastery.com/index-slice-reshape-numpy-arrays-machine-learning-python/
- https://machinelearningmastery.com/broadcasting-with-numpy-arrays/

and make notes on

1) Index, Slice and Reshape NumPy Arrays
2) NumPy Array Broadcasting

- You can download the "Free mini courses" by providing your email address

Further Reading

- Check the following URLS
- https://machinelearningmastery.com/introduction-matrices-machine-learning/
- https://machinelearningmastery.com/matrix-operations-for-machine-learning/
- https://machinelearningmastery.com/introduction-to-matrix-decompositions-for-machine-learning/

and make notes on

1. Matrices: Vectors and Vector Arithmetic, Vector Norms, Matrices and Matrix Arithmetic, Types of matrices, matrix operations, sparse matrices, Tensors and tensor arithmetic
2. Factorization: Matrix decompositions, eigendecomposition, singular value decomposition

- You can download the “Free mini courses” by providing your email address

The End

Install Python 3.6 in your laptops.

See .pdf file titled “Python Ecosystem for Machine Learning”