Deep Learning from Ground Up

From Linear Algebra to Statistics

Akshay Badola
15MCPC15

School of Computer and Information Sciences
University of Hyderabad

December 2017

Overview

Linear Algebra

Why Linear Algebra?

- What's an algebra?

Linear Algebra

Why Linear Algebra?

- What's an algebra? A Mathematical system for doing computations.

Linear Algebra

Why Linear Algebra?

- What's an algebra? A Mathematical system for doing computations.
- Isn't that just all of mathematics?

Linear Algebra

Why Linear Algebra?

- What's an algebra? A Mathematical system for doing computations.
- Isn't that just all of mathematics? No, and what it is, is beyond the scope of this talk!

Linear Algebra

Why Linear Algebra?

- What's an algebra? A Mathematical system for doing computations.
- Isn't that just all of mathematics? No, and what it is, is beyond the scope of this talk!
- So, it's a system for doing computations and it has two key properties:
  - Linear
  - Algebra
- By linear we mean that the constituents of the system can be expressed simply by counting.
- Algebra reflects certain properties of the system, so that the computations remain consistent so to speak.
- It is assumed that you've studied some Linear Algebra already and are familiar with Matrices, Determinants, Inverses, Adjoints at the very least.
- It's not going to be thorough introduction, but should clear some things.

Formally I

A bit more...

- We'll deal with the notion of Linear (or Vector) Spaces.
- Formally a Linear Space over a **field** F (often the field of the real numbers but can be complex numbers for our purposes) is a set V equipped with two binary operations satisfying the following axioms.
- Elements of V are called *vectors*, and elements of F are called *scalars*. The first operation, vector addition, takes any two vectors **v** and **v** and outputs a third vector $\mathbf{v} + \mathbf{w}$.
- The second operation, scalar multiplication, takes any scalar a and any vector **v** and outputs a new vector $a\mathbf{v}$.
- The operations of addition and multiplication in a vector space must satisfy the following axioms.

  In the list below, let **u, v** and **w** be arbitrary vectors in V, and a and b scalars in F.
    - Associativity of addition: $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$

Formally II

- Commutativity of addition: $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$
- Identity element of addition: There exists an element $\mathbf{0} \in V$, called the zero vector, such that $\mathbf{v} + \mathbf{0} = \mathbf{v}$ for all $\mathbf{v} \in V$.
- Inverse elements of addition: For every $\mathbf{v} \in V$, there exists an element $-\mathbf{v} \in V$, called the additive inverse of $\mathbf{v}$, such that $\mathbf{v} + (-\mathbf{v}) = 0$
- Distributivity of scalar multiplication with respect to vector addition: $a(\mathbf{u} + \mathbf{v}) = a\mathbf{u} + a\mathbf{v}$
- Distributivity of scalar multiplication with respect to field addition: $(a + b)\mathbf{v} = a\mathbf{v} + b\mathbf{v}$
- Compatibility of scalar multiplication with field multiplication: $a(b\mathbf{v}) = (ab)\mathbf{v}$
- Identity element of scalar multiplication: $1\mathbf{v} = \mathbf{v}$, where 1 denotes the multiplicative identity in F.

Why Linear Algebra?

Because it's the only way there is

- For humans (and computers) recursion is a natural way to think (for some unfathomable reason).
- Any mathematical system starts with the number 0 and defines 1 simply as the next number.
- 2 can be defined simply by addition of 1 to 1 which is nothing but counting one step ahead and so on for the rest.

Why Linear Algebra?

Because it's the only way there is

- For humans (and computers) recursion is a natural way to think (for some unfathomable reason).
- Any mathematical system starts with the number 0 and defines 1 simply as the next number.
- 2 can be defined simply by addition of 1 to 1 which is nothing but counting one step ahead and so on for the rest.
- The notion that any number can be arrived at, by taking some countable number of steps is a fundamental way in which we think.
- Linear Spaces is just a formalization of that.
- How?

One important note: We'll assume throughout that all vectors are column vectors, so that $\mathbf{u}$ can be imagined as a column of scalars, so that $\mathbf{Au}$ again produces a column of scalars, while $\mathbf{uv}^T$ produces a matrix.

Why Linear Algebra?

Even a child can count!

- How?
  - Any vector's components can be counted. (Remember, counting is equivalent to addition)

Why Linear Algebra?

Even a child can count!

- How?
  - Any vector's components can be counted. (Remember, counting is equivalent to addition)
  - Two vectors can always be combined in a countable way, and thus by extension, any countable number of vectors can be combined in a countable way.

Why Linear Algebra?

Even a child can count!

- How?
  - Any vector's components can be counted. (Remember, counting is equivalent to addition)
  - Two vectors can always be combined in a countable way, and thus by extension, any countable number of vectors can be combined in a countable way.
  - Multiplication of two vectors (scalar product) involves also doing countable computations, as their components are multiplied correspondingly. If the number of components differ, then they can't be multiplied.

Why Linear Algebra?

Even a child can count!

- How?
  - Any vector's components can be counted. (Remember, counting is equivalent to addition)
  - Two vectors can always be combined in a countable way, and thus by extension, any countable number of vectors can be combined in a countable way.
  - Multiplication of two vectors (scalar product) involves also doing countable computations, as their components are multiplied correspondingly. If the number of components differ, then they can't be multiplied.
  - And these are in essence all of the operations we need to get any other vector from that space. Right?

Why Linear Algebra? I

Because we need something to represent the world

- Assignment [1]

  An **assignment on** I or $I-$ **assignment**, $f = (f_i)_{i \in I} = (f_i : i \in I)$ associates with each element i in its **domain** some **term** or **item** or **value** f_i
  $f:$ dom $f : i \rightarrow f_i,$ the set ran $f := \{f_i : i \in$ dom $f\}$

  > *During the Cholera outbreak in 1854 in London, Dr. John Snow recorded the deaths by address, thus setting up an assignment whose domain consisted of all the houses in London. But he did not simply make a list of all the addresses and then record the deaths in that list. Rather, he took a map of London and marked the number of deaths at each address right on the map (not bothering to record the value 0 of no death). He found that the deaths clustered around one particular public water pump, jumped to a conclusion (remember that this was well before Pasteurs discoveries), had the handle of that pump removed and had the satisfaction of seeing the epidemic fade.*

Why Linear Algebra? II

> *Thus, one way to think of an assignment is to visualize its domain in some convenient fashion, and, at each element of the domain, its assigned item or value. This is routinely done for matrices, another basic object in these notes. [1]*

- So essentially a vector space can be considered as a map from objects of the real world, to the set of numbers where the numbers are certain properties of those objects.
- In order to *quantify* the real world, this is the only way we can do it.

A Different View

Inferring the missing values

- Matrices had originally been developed to solve systems of linear equations.
- As the name suggests, they are multiple equations where each variable corresponds to a component of a Linear Space and the coefficients correspond to the scalars.

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ \vdots & & \vdots & & & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array}$$

A Different View

Inferring the missing values

- Matrices had originally been developed to solve systems of linear equations.
- As the name suggests, they are multiple equations where each variable corresponds to a component of a Linear Space and the coefficients correspond to the scalars.

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ \vdots & & \vdots & & & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array}$$

- This can be more compactly represented as:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \ddots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

A Different View

Inferring the missing values

- Matrices had originally been developed to solve systems of linear equations.
- As the name suggests, they are multiple equations where each variable corresponds to a component of a Linear Space and the coefficients correspond to the scalars.

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ \vdots & & \vdots & & & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array}$$

- This can be more compactly represented as:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

- Or $\mathbf{Ax} = \mathbf{b}$

Matrix Multiplication

Have you ever wondered why are matrices multiplied like they are and not simply elementwise?

Matrix Multiplication

Have you ever wondered why are matrices multiplied like they are and not simply elementwise?

- Well they can be multiplied elementwise also, which is called a **Hadamard Product**.
- But when we talk about Matrix Multiplication we mean a **Row to Column** correspondence. Why's it like that?

Matrix Multiplication

Have you ever wondered why are matrices multiplied like they are and not simply elementwise?

- Well they can be multiplied elementwise also, which is called a **Hadamard Product**.
- But when we talk about Matrix Multiplication we mean a **Row to Column** correspondence. Why's it like that?
- Let's take a closer look at Matrix/Vector Multiplication.
  For $\mathbf{Ax} = \mathbf{b}, \mathbf{x} \in \mathcal{R}^n, \mathbf{b} \in \mathcal{R}^m$
  The matrix $\mathbf{A}$ takes a vector of n components and converts it to one of m components.

Matrix Multiplication

Have you ever wondered why are matrices multiplied like they are and not simply elementwise?

- Well they can be multiplied elementwise also, which is called a **Hadamard Product**.
- But when we talk about Matrix Multiplication we mean a **Row to Column** correspondence. Why's it like that?
- Let's take a closer look at Matrix/Vector Multiplication.
  For $\mathbf{Ax} = \mathbf{b}, \mathbf{x} \in \mathcal{R}^n, \mathbf{b} \in \mathcal{R}^m$
  The matrix **A** takes a vector of *n* components and converts it to one of *m* components.
- Assuming a utilitarian viewpoint, for the moment we assume that the vectors *assignments* from the real world as mentioned earlier and represent some properties of objects of interest to us.
- Also, let's say we want to find a way to relate these quantities to some other set of quantities, but with a different number of components. How would you do it?

Matrix Multiplication

Formalizing it...

- Formalizing it, $\mathbf{Ax}_1 = \mathbf{b}_1, \mathbf{Ax}_2 = \mathbf{b}_2, ..., \mathbf{Ax}_k = \mathbf{b}_k$, or $\mathbf{AX} = \mathbf{B}$, where $\mathbf{A} \in \mathcal{R}^{n,m}, \mathbf{X} \in \mathcal{R}^{m,k}, \mathbf{B} \in \mathcal{R}^{n,k}$

Matrix Multiplication

Formalizing it...

- Formalizing it, $\mathbf{Ax}_1 = \mathbf{b}_1, \mathbf{Ax}_2 = \mathbf{b}_2, ..., \mathbf{Ax}_k = \mathbf{b}_k$, or

  $\mathbf{AX} = \mathbf{B}$, where $\mathbf{A} \in \mathcal{R}^{n,m}, \mathbf{X} \in \mathcal{R}^{m,k}, \mathbf{B} \in \mathcal{R}^{n,k}$
- So, informally $\mathbf{A}$ defines a sort of association or a rule which maps some data with n components to some other data with m components.

Matrix Multiplication

Formalizing it...

- Formalizing it, $\mathbf{A}\mathbf{x}_1 = \mathbf{b}_1, \mathbf{A}\mathbf{x}_2 = \mathbf{b}_2, ..., \mathbf{A}\mathbf{x}_k = \mathbf{b}_k$, or

  $\mathbf{AX} = \mathbf{B}$, where $\mathbf{A} \in \mathcal{R}^{n,m}, \mathbf{X} \in \mathcal{R}^{m,k}, \mathbf{B} \in \mathcal{R}^{n,k}$
- So, informally $\mathbf{A}$ defines a sort of association or a rule which maps some data with n components to some other data with m components.
- Formally, we say that $\mathbf{A}$ is a *Linear Map* or a *Linear Transformation* from space $\mathcal{R}^n \rightarrow \mathcal{R}^m$.

Matrix Multiplication

Formalizing it...

- Formalizing it, $\mathbf{Ax}_1 = \mathbf{b}_1, \mathbf{Ax}_2 = \mathbf{b}_2, ..., \mathbf{Ax}_k = \mathbf{b}_k$, or

  $\mathbf{AX = B}$, where $\mathbf{A} \in \mathcal{R}^{n,m}, \mathbf{X} \in \mathcal{R}^{m,k}, \mathbf{B} \in \mathcal{R}^{n,k}$
- So, informally **A** defines a sort of association or a rule which maps some data with n components to some other data with m components.
- Formally, we say that **A** is a *Linear Map* or a *Linear Transformation* from space $\mathcal{R}^n \rightarrow \mathcal{R}^m$.
- Now what is this Linear Map doing? Also it looks so much like a function, right?

Matrix Multiplication

Formalizing it...

- Formalizing it, $\mathbf{Ax}_1 = \mathbf{b}_1, \mathbf{Ax}_2 = \mathbf{b}_2, ..., \mathbf{Ax}_k = \mathbf{b}_k$, or

  $\mathbf{AX} = \mathbf{B}$, where $\mathbf{A} \in \mathcal{R}^{n,m}, \mathbf{X} \in \mathcal{R}^{m,k}, \mathbf{B} \in \mathcal{R}^{n,k}$
- So, informally **A** defines a sort of association or a rule which maps some data with n components to some other data with m components.
- Formally, we say that **A** is a *Linear Map* or a *Linear Transformation* from space $\mathcal{R}^n \rightarrow \mathcal{R}^m$.
- Now what is this Linear Map doing? Also it looks so much like a function, right?
- Let's say for $\mathcal{R}^n \rightarrow \mathcal{R}^m$, this map **A** is not the Matrix Multiplication as we see it, but some other Linear Operation. What can it be?

Matrix Multiplication

Formalizing it...

- Formalizing it, $\mathbf{Ax}_1 = \mathbf{b}_1, \mathbf{Ax}_2 = \mathbf{b}_2, ..., \mathbf{Ax}_k = \mathbf{b}_k$, or

  $\mathbf{AX} = \mathbf{B}$, where $\mathbf{A} \in \mathcal{R}^{n,m}, \mathbf{X} \in \mathcal{R}^{m,k}, \mathbf{B} \in \mathcal{R}^{n,k}$
- So, informally **A** defines a sort of association or a rule which maps some data with n components to some other data with m components.
- Formally, we say that **A** is a *Linear Map* or a *Linear Transformation* from space $\mathcal{R}^n \to \mathcal{R}^m$.
- Now what is this Linear Map doing? Also it looks so much like a function, right?
- Let's say for $\mathcal{R}^n \to \mathcal{R}^m$, this map **A** is not the Matrix Multiplication as we see it, but some other Linear Operation. What can it be?
- Find the most important scalars? But how to find them? Even if we did, is this Linear?
- Pick random scalars from **X**, with replacement until you have enough scalars? Is this Linear?
- How about take a weighted average of the data with n components? Is this Linear?

Matrix Multiplication

Eureka!

- First may be linear or may not be, depending on the method used to find the *important components*. Second is not linear. Last is exactly Matrix Multiplication.
- Notice that for the set of observations $\mathbf{X}$, each column of $\mathbf{A}$ multiplies with $\mathbf{x}_i$ and contributes to the corresponding scalar of $\mathbf{b}_i$ where $\mathbf{x}_i, \mathbf{b}_i$ is any arbitrary pair from $\mathbf{X}$, $\mathbf{B}$.

Matrix Multiplication

Eureka!

- First may be linear or may not be, depending on the method used to find the *important components*. Second is not linear. Last is exactly Matrix Multiplication.
- Notice that for the set of observations $\mathbf{X}$, each column of $\mathbf{A}$ multiplies with $\mathbf{x}_i$ and contributes to the corresponding scalar of $\mathbf{b}_i$ where $\mathbf{x}_i, \mathbf{b}_i$ is any arbitrary pair from $\mathbf{X}$, $\mathbf{B}$.
- Because $\mathbf{x}_i.\mathbf{a}_j$ is the dot product between the vector $\mathbf{x}_i$ and j^{th} column $\mathbf{A}$.
- In a sense, each component of $\mathbf{b}_i$ contains some contribution from each component of $\mathbf{x}_i$, multiplied by some weights in $\mathbf{A}$, which is exactly what you'd want!

Matrix Multiplication

Eureka!

- First may be linear or may not be, depending on the method used to find the *important components*. Second is not linear. Last is exactly Matrix Multiplication.
- Notice that for the set of observations $\mathbf{X}$, each column of $\mathbf{A}$ multiplies with $\mathbf{x}_i$ and contributes to the corresponding scalar of $\mathbf{b}_i$ where $\mathbf{x}_i, \mathbf{b}_i$ is any arbitrary pair from $\mathbf{X}$, $\mathbf{B}$.
- Because $\mathbf{x}_i.\mathbf{a}_j$ is the dot product between the vector $\mathbf{x}_i$ and j^{th} column $\mathbf{A}$.
- In a sense, each component of $\mathbf{b}_i$ contains some contribution from each component of $\mathbf{x}_i$, multiplied by some weights in $\mathbf{A}$, which is exactly what you'd want! And it's Linear!
- Henceforth, we'll call the number of components in both the Linear Transformation $\mathbf{A}$ and the vector $\mathbf{x}$ as their *dimension*. Though it's a much more complicated and deeper concept.
- We'll also assume that any dimension we talk of is finite unless otherwise stated.

Normed Linear Spaces

What's a vector worth?

- So hopefully by now, we understand the need and constructs of Linear Spaces. But even though we can now construct arbitrary Linear Spaces and Maps, we still have something missing. What is it?

Normed Linear Spaces

What's a vector worth?

- So hopefully by now, we understand the need and constructs of Linear Spaces. But even though we can now construct arbitrary Linear Spaces and Maps, we still have something missing. What is it?
- A notion of a value. Which brings with the notion of order. We want to know whether a vector has greater value than the other, we want to know if they have the same value.
- For instance we're sure that the vector $< 0, 0, 0 >$ has less value than $< 1, 1, 1 >$, but how do we quantify it?

Normed Linear Spaces

What's a vector worth?

- So hopefully by now, we understand the need and constructs of Linear Spaces. But even though we can now construct arbitrary Linear Spaces and Maps, we still have something missing. What is it?
- A notion of a value. Which brings with the notion of order. We want to know whether a vector has greater value than the other, we want to know if they have the same value.
- For instance we're sure that the vector $< 0, 0, 0 >$ has less value than $< 1, 1, 1 >$, but how do we quantify it?
- For that we define the concept of a *Norm*. The Norm gives us the value of any element in a computational system. Formally speaking, for Linear Spaces, a Norm is a function $p : V \rightarrow R$ with the following properties:
  - $p(a\mathbf{v}) = |a|p(\mathbf{v})$. Homogenous.
  - $p(\mathbf{v}) \geq 0$. Non-negative.
  - If $p(\mathbf{v}) = 0$ then $\mathbf{v} = \mathbf{0}$. Is 0 only for the zero or **null** vector.
  - $p(\mathbf{u} + \mathbf{v}) \leq p(\mathbf{u}) + p(\mathbf{v})$. Satisfies the triangle inequality.

Distance in Normed Linear Spaces

What's in a Norm?

- There may be Linear Spaces without a Norm. But spaces with a Normed Spaces some special qualities which are desirable and make them easy to work with. We won't speak of spaces without norms.
- For a Linear Space what can be a Norm? Can there be multiple Norms? Are they equivalent?

Distance in Normed Linear Spaces

What's in a Norm?

- There may be Linear Spaces without a Norm. But spaces with a Normed Spaces some special qualities which are desirable and make them easy to work with. We won't speak of spaces without norms.
- For a Linear Space what can be a Norm? Can there be multiple Norms? Are they equivalent?
- For a vector **x**, the absolute value, $\sum |x_i|$ is a norm. $\sum {x_i}^2$ is a norm, etc. There can be many different norms and we won't go into the details here. For most of the cases we'll deal with the Euclidean Norm or $\sqrt{\sum {x_i}^2}$.
- Note that $\sqrt{\sum {x_i}^2}$ and $\sum {x_i}^2$ are equivalent except for a scaling factor.

Distance in Normed Linear Spaces

What's in a Norm?

- There may be Linear Spaces without a Norm. But spaces with a Normed Spaces some special qualities which are desirable and make them easy to work with. We won't speak of spaces without norms.
- For a Linear Space what can be a Norm? Can there be multiple Norms? Are they equivalent?
- For a vector **x**, the absolute value, $\sum |x_i|$ is a norm. $\sum {x_i}^2$ is a norm, etc. There can be many different norms and we won't go into the details here. For most of the cases we'll deal with the Euclidean Norm or $\sqrt{\sum {x_i}^2}$.
- Note that $\sqrt{\sum {x_i}^2}$ and $\sum {x_i}^2$ are equivalent except for a scaling factor.
- What norm does is bring a notion of distance between two vectors, since they each have now a notion of a value, their values can be compared.
- The distance is called the *metric* induced by the norm. In case of the standard Euclidean norm for two vectors $\mathbf{x}_1, \mathbf{x}_2$ it is, $\sqrt{\sum (x_{1,i} - x_{2,i})^2}$

The Angle

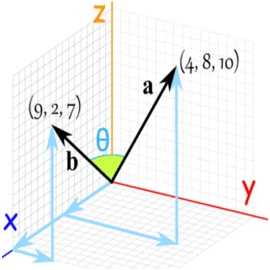
The Angle implies a Direction

- However the Vector Spaces we'll be talking about will have one additional quality: a notion of an inner product.
- The inner product is the *dot product* with which you're all familiar. What it confers to the vectors is a *direction*

The Angle

The Angle implies a Direction

- However the Vector Spaces we'll be talking about will have one additional quality: a notion of an inner product.
- The inner product is the *dot product* with which you're all familiar. What it confers to the vectors is a *direction*
- The cosine of the angle between two vectors is defined as $\frac{\mathbf{x}_1.\mathbf{x}_2}{|\mathbf{x}_1||\mathbf{x}_2|}$
- As we'll later see, it can be seen as a measure of similarity between two vectors.



More on Linear Transformations

Linear Dependence and Independence

- We've seen that a Matrix is a Linear Map from one vector space to another.
- But from which space to which remains to be ascertained.
- Recall that a Matrix is simply the coefficients for a system of linear equations. Let's consider the following set of equations

$$
\begin{aligned}
3x_1 + 4x_2 + 5x_3 = 8 \\
3x_1 + 4x_2 + 5x_3 = 9 \\
3x_1 + 4x_2 + 5x_3 = 10
\end{aligned}
$$

More on Linear Transformations

Linear Dependence and Independence

- We've seen that a Matrix is a Linear Map from one vector space to another.
- But from which space to which remains to be ascertained.
- Recall that a Matrix is simply the coefficients for a system of linear equations. Let's consider the following set of equations

$$\begin{aligned} 3x_1 + 4x_2 + 5x_3 &= 8 \\ 3x_1 + 4x_2 + 5x_3 &= 9 \\ 3x_1 + 4x_2 + 5x_3 &= 10 \end{aligned}$$

- Again, it can be more conveniently represented as:

$$\begin{bmatrix} 3 & 4 & 5 \\ 3 & 4 & 5 \\ 3 & 4 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 8 \\ 9 \\ 10 \end{bmatrix}$$

More on Linear Transformations

Linear Dependence and Independence

- We've seen that a Matrix is a Linear Map from one vector space to another.
- But from which space to which remains to be ascertained.
- Recall that a Matrix is simply the coefficients for a system of linear equations. Let's consider the following set of equations

$$
\begin{aligned}
3x_1 + 4x_2 + 5x_3 = 8 \\
3x_1 + 4x_2 + 5x_3 = 9 \\
3x_1 + 4x_2 + 5x_3 = 10
\end{aligned}
$$

- Again, it can be more conveniently represented as:

$$
\begin{bmatrix} 3 & 4 & 5 \\ 3 & 4 & 5 \\ 3 & 4 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 8 \\ 9 \\ 10 \end{bmatrix}
$$

- But there seems to be something amiss in the formulation above.

More on Linear Transformations

Linear Dependence and Independence

- We've seen that a Matrix is a Linear Map from one vector space to another.
- But from which space to which remains to be ascertained.
- Recall that a Matrix is simply the coefficients for a system of linear equations. Let's consider the following set of equations

$$\begin{aligned} 3x_1 + 4x_2 + 5x_3 = 8 \\ 3x_1 + 4x_2 + 5x_3 = 9 \\ 3x_1 + 4x_2 + 5x_3 = 10 \end{aligned}$$

- Again, it can be more conveniently represented as:

$$\begin{bmatrix} 3 & 4 & 5 \\ 3 & 4 & 5 \\ 3 & 4 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 8 \\ 9 \\ 10 \end{bmatrix}$$

- But there seems to be something amiss in the formulation above.
- These equations can't be solved! There's no vector **x** which'll map to **b** in that case.
- This system is called *inconsistent*.

Linear Dependence and Independence

Linear Dependence and Independence

- What about these?

$$\begin{bmatrix} 3 & 4 & 5 \\ 6 & 8 & 10 \\ 8 & 10 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 8 \\ 16 \\ 12 \end{bmatrix}$$

Linear Dependence and Independence

Linear Dependence and Independence

- What about these?

$$\begin{bmatrix} 3 & 4 & 5 \\ 6 & 8 & 10 \\ 8 & 10 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 8 \\ 16 \\ 12 \end{bmatrix}$$

- And these?

$$\begin{bmatrix} 3 & 4 & 5 \\ 3 & 8 & 6 \\ 8 & 3 & 3 \\ 1 & 2 & 6 \\ 9 & 4 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 8 \\ 16 \\ 4 \\ 0 \\ 12 \end{bmatrix}$$

Linear Dependence and Independence

Linear Dependence and Independence

- What about these?

$$\begin{bmatrix} 3 & 4 & 5 \\ 6 & 8 & 10 \\ 8 & 10 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 8 \\ 16 \\ 12 \end{bmatrix}$$

- And these?

$$\begin{bmatrix} 3 & 4 & 5 \\ 3 & 8 & 6 \\ 8 & 3 & 3 \\ 1 & 2 & 6 \\ 9 & 4 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 8 \\ 16 \\ 4 \\ 0 \\ 12 \end{bmatrix}$$

- The last system is called *overdetermined* and usually has no solution. Previous to that is underdetermined and has infinite solutions.

Linear Dependence and Independence

Linear Dependence and Independence

- What about these?

$$\begin{bmatrix} 3 & 4 & 5 \\ 6 & 8 & 10 \\ 8 & 10 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 8 \\ 16 \\ 12 \end{bmatrix}$$

- And these?

$$\begin{bmatrix} 3 & 4 & 5 \\ 3 & 8 & 6 \\ 8 & 3 & 3 \\ 1 & 2 & 6 \\ 9 & 4 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 8 \\ 16 \\ 4 \\ 0 \\ 12 \end{bmatrix}$$

- The last system is called *overdetermined* and usually has no solution. Previous to that is underdetermined and has infinite solutions.
- An overdetermined system can only be solved if some of the equations are linear combination of others.

Linear Dependence and Independence

Linear Dependence and Independence

- A linear combination is simply $a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + a_3\mathbf{x}_3 ... a_n\mathbf{x}_n$
- If for some linear combination $a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + a_3\mathbf{x}_3 ... a_n\mathbf{x}_n = \mathbf{x}_k$, then we say that the vector $\mathbf{x}_k$ is a linear combination of the vectors on LHS.
- If for some linear combination $a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + a_3\mathbf{x}_3 ... a_n\mathbf{x}_n = \mathbf{0}$, where $\mathbf{0}$ is the null vector, (a_i not all zero), then we say that the vectors are *linearly independent*.

$$\begin{bmatrix} 3 & 4 & 5 \\ 6 & 4 & 10 \\ 8 & 10 & 10 \\ 12 & 12 & 20 \\ 17 & 18 & 25 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 8 \\ 8 \\ 12 \\ 24 \\ 28 \end{bmatrix}$$

- What bout the above equation, can it be solved?

Linear Dependence and Independence

Linear Dependence and Independence

- A linear combination is simply $a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + a_3\mathbf{x}_3 ... a_n\mathbf{x}_n$
- If for some linear combination $a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + a_3\mathbf{x}_3 ... a_n\mathbf{x}_n = \mathbf{x}_k$, then we say that the vector $\mathbf{x}_k$ is a linear combination of the vectors on LHS.
- If for some linear combination $a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + a_3\mathbf{x}_3 ... a_n\mathbf{x}_n = \mathbf{0}$, where $\mathbf{0}$ is the null vector, (a_i not all zero), then we say that the vectors are *linearly independent*.

$$\begin{bmatrix} 3 & 4 & 5 \\ 6 & 4 & 10 \\ 8 & 10 & 10 \\ 12 & 12 & 20 \\ 17 & 18 & 25 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 8 \\ 8 \\ 12 \\ 24 \\ 28 \end{bmatrix}$$

- What bout the above equation, can it be solved? Yes! Because two of the equations are linear combinations of others.

Linear Dependence and Independence

Linear Dependence and Independence

- A linear combination is simply $a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + a_3\mathbf{x}_3 ... a_n\mathbf{x}_n$
- If for some linear combination $a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + a_3\mathbf{x}_3 ... a_n\mathbf{x}_n = \mathbf{x}_k$, then we say that the vector $\mathbf{x}_k$ is a linear combination of the vectors on LHS.
- If for some linear combination $a_1\mathbf{x}_1 + a_2\mathbf{x}_2 + a_3\mathbf{x}_3 ... a_n\mathbf{x}_n = \mathbf{0}$, where $\mathbf{0}$ is the null vector, (a_i not all zero), then we say that the vectors are *linearly independent*.

$$\begin{bmatrix} 3 & 4 & 5 \\ 6 & 4 & 10 \\ 8 & 10 & 10 \\ 12 & 12 & 20 \\ 17 & 18 & 25 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 8 \\ 8 \\ 12 \\ 24 \\ 28 \end{bmatrix}$$

- What bout the above equation, can it be solved? Yes! Because two of the equations are linear combinations of others.
- What this may mean in one sense, is that they supply no extra information, or may be redundant.

Linear Dependence and Independence

Linear Dependence and Independence

- Note that for all the matrices, whatever be the value of the (let's say input) vector **x** that is multiplied, there exists some (output) vector **b**.

$$\begin{bmatrix} 3 & 4 & 5 \\ 3 & 4 & 5 \\ 3 & 4 & 5 \end{bmatrix} \begin{bmatrix} 3 & 4 & 5 \\ 6 & 8 & 10 \\ 8 & 10 & 10 \end{bmatrix} \begin{bmatrix} 3 & 4 & 5 \\ 3 & 8 & 6 \\ 8 & 3 & 3 \\ 1 & 2 & 6 \\ 9 & 4 & 3 \end{bmatrix} \begin{bmatrix} 3 & 4 & 5 \\ 6 & 4 & 10 \\ 8 & 10 & 10 \\ 12 & 12 & 20 \\ 17 & 18 & 25 \end{bmatrix} \begin{bmatrix} 2 & 3 & 4 & 5 \\ 5 & 6 & 4 & 10 \\ 1 & 8 & 10 & 10 \end{bmatrix}$$

- The last Matrix defines a 4x3 transformation. But if there are some dependent rows (or columns) in there, then perhaps the output vector may be smaller.
- The fact that only the number of *independent* rows or columns matters on the output vector is essential in a lot of Linear Algebra and Statistics.

Linear Dependence and Independence

Linear Dependence and Independence

- Note that for all the matrices, whatever be the value of the (let's say input) vector **x** that is multiplied, there exists some (output) vector **b**.

$$\begin{bmatrix} 3 & 4 & 5 \\ 3 & 4 & 5 \\ 3 & 4 & 5 \end{bmatrix} \begin{bmatrix} 3 & 4 & 5 \\ 6 & 8 & 10 \\ 8 & 10 & 10 \end{bmatrix} \begin{bmatrix} 3 & 4 & 5 \\ 3 & 8 & 6 \\ 8 & 3 & 3 \\ 1 & 2 & 6 \\ 9 & 4 & 3 \end{bmatrix} \begin{bmatrix} 3 & 4 & 5 \\ 6 & 4 & 10 \\ 8 & 10 & 10 \\ 12 & 12 & 20 \\ 17 & 18 & 25 \end{bmatrix} \begin{bmatrix} 2 & 3 & 4 & 5 \\ 5 & 6 & 4 & 10 \\ 1 & 8 & 10 & 10 \end{bmatrix}$$

- The last Matrix defines a 4x3 transformation. But if there are some dependent rows (or columns) in there, then perhaps the output vector may be smaller.
- The fact that only the number of *independent* rows or columns matters on the output vector is essential in a lot of Linear Algebra and Statistics.
- So the fact that even though the output vector appears to be of dimension 10, it may only be that most of the components are being determined by a few independent vectors in the transformation. So the effective dimension may be less!

Linear Dependence and Independence

Linear Spaces and Independence

- Let's say you have been given pairs of vectors and it's told that there exists a *Linear* relation between those pairs. Say, these arrays:

$$\begin{bmatrix} 3 & 4 & 5 \\ 6 & 4 & 10 \\ 8 & 10 & 10 \\ 12 & 12 & 20 \\ 17 & 18 & 25 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 1 & 1 \\ 3 & 5 \\ 2 & 6 \\ 4 & 3 \end{bmatrix}$$

- Clearly it would be a 3x2 Matrix. And we don't know the values of the elements of the matrix.
- And if some of these pairs are linear combinations of others, then they add no information.
- And if some columns in the data on LHS are linear combinations of other columns, then they can be eliminated.

Linear Dependence and Independence

Linear Spaces and Independence

- Let's say you have been given pairs of vectors and it's told that there exists a *Linear* relation between those pairs. Say, these arrays:

$$\begin{bmatrix} 3 & 4 & 5 \\ 6 & 4 & 10 \\ 8 & 10 & 10 \\ 12 & 12 & 20 \\ 17 & 18 & 25 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 1 & 1 \\ 3 & 5 \\ 2 & 6 \\ 4 & 3 \end{bmatrix}$$

- Clearly it would be a 3x2 Matrix. And we don't know the values of the elements of the matrix.
- And if some of these pairs are linear combinations of others, then they add no information.
- And if some columns in the data on LHS are linear combinations of other columns, then they can be eliminated.

Linear Dependence and Independence

Linear Spaces and Independence

- Let's say you have been given pairs of vectors and it's told that there exists a *Linear* relation between those pairs. Say, these arrays:

$$\begin{bmatrix} 3 & 4 & 5 \\ 6 & 4 & 10 \\ 8 & 10 & 10 \\ 12 & 12 & 20 \\ 17 & 18 & 25 \end{bmatrix}, \begin{bmatrix} 1 & 2 \\ 1 & 1 \\ 3 & 5 \\ 2 & 6 \\ 4 & 3 \end{bmatrix}$$

- Clearly it would be a 3x2 Matrix. And we don't know the values of the elements of the matrix.
- And if some of these pairs are linear combinations of others, then they add no information.
- And if some columns in the data on LHS are linear combinations of other columns, then they can be eliminated.

Switching to Vector Spaces

- A set of vectors *spans* a Linear (Vector) Space, if all of the vectors in that space can be obtained from linear combinations in that set.
- That set is called a basis, if it is the smallest such set.
- The cardinality of a basis (there can be multiple bases) is called the

Linear Dependence and Independence

Dimensionality

- Every vector space has a basis of course. The natural basis is $\{0, 0, 0, 0...1...0, 0, 0\}$, with 1 at each position and 0_s in the rest. Clearly it spans the entire space.

Linear Dependence and Independence

Dimensionality

- Every vector space has a basis of course. The natural basis is $\{0, 0, 0, 0...1...0, 0, 0\}$, with 1 at each position and 0_s in the rest. Clearly it spans the entire space.
- Coming back to the matrices in the previous slide: imagine if the data size was very large, i.e., the number of rows and columns on both LHS and RHS were really large, it would clearly help to find the real dimension of the data instead of the very large one that is evident.
- The corresponding Linear Transformation (or matrix) would also be smaller.

Linear Dependence and Independence

Dimensionality

- Every vector space has a basis of course. The natural basis is $\{0, 0, 0, 0...1...0, 0, 0\}$, with 1 at each position and 0_s in the rest. Clearly it spans the entire space.
- Coming back to the matrices in the previous slide: imagine if the data size was very large, i.e., the number of rows and columns on both LHS and RHS were really large, it would clearly help to find the real dimension of the data instead of the very large one that is evident.
- The corresponding Linear Transformation (or matrix) would also be smaller.

We'll review one more properties of Linear Spaces and their Transformations, which is Orthogonality and then we'll go into Eigenvalues and Eigenvectors.

Orthogonality

Some vectors matter more

- From the earlier discussions, it can be seen clearly that to describe a vector space, it's enough to know the basis. But what can be an admissible basis?

Orthogonality

Some vectors matter more

- From the earlier discussions, it can be seen clearly that to describe a vector space, it's enough to know the basis. But what can be an admissible basis?
- We define here the concept of *Orthogonal* Vectors and Matrices. Two vectors are said to be orthogonal if their inner product is zero, i.e., $\mathbf{x}_1.\mathbf{x}_2 = 0$
- A Matrix is said to be orthogonal if its product with its transpose is the identity matrix, i.e., $\mathbf{A}\mathbf{A}^T = \mathbf{I}$. (Unitary if $\mathbf{F}$ is complex)
- The notion of direction is very important here, as two orthogonal vectors are perpendicular to each other. But what about the Matrix?

Orthogonality

Some vectors matter more

- From the earlier discussions, it can be seen clearly that to describe a vector space, it's enough to know the basis. But what can be an admissible basis?
- We define here the concept of *Orthogonal* Vectors and Matrices. Two vectors are said to be orthogonal if their inner product is zero, i.e., $\mathbf{x}_1.\mathbf{x}_2 = 0$
- A Matrix is said to be orthogonal if its product with its transpose is the identity matrix, i.e., $\mathbf{A}\mathbf{A}^T = \mathbf{I}$. (Unitary if $\mathbf{F}$ is complex)
- The notion of direction is very important here, as two orthogonal vectors are perpendicular to each other. But what about the Matrix?
- The Matrix $\mathbf{A}$ which is orthogonal, defines a linear transformation that preserves direction. To see that, observe:
  For, $\mathbf{u}, \mathbf{v} \in R^n, \mathbf{Au}.\mathbf{Av} = (\mathbf{uA})^T.\mathbf{Av} = \mathbf{u}^T(\mathbf{A}^T\mathbf{A})\mathbf{v} = \mathbf{u}^T\mathbf{v} = \mathbf{u}.\mathbf{v}$

Orthogonality

Some vectors matter more

- From the earlier discussions, it can be seen clearly that to describe a vector space, it's enough to know the basis. But what can be an admissible basis?
- We define here the concept of *Orthogonal* Vectors and Matrices. Two vectors are said to be orthogonal if their inner product is zero, i.e., $\mathbf{x}_1.\mathbf{x}_2 = 0$
- A Matrix is said to be orthogonal if its product with its transpose is the identity matrix, i.e., $\mathbf{A}\mathbf{A}^T = \mathbf{I}$. (Unitary if $\mathbf{F}$ is complex)
- The notion of direction is very important here, as two orthogonal vectors are perpendicular to each other. But what about the Matrix?
- The Matrix $\mathbf{A}$ which is orthogonal, defines a linear transformation that preserves direction. To see that, observe:
  For, $\mathbf{u}, \mathbf{v} \in R^n, \mathbf{Au}.\mathbf{Av} = (\mathbf{uA})^T.\mathbf{Av} = \mathbf{u}^T(\mathbf{A}^T\mathbf{A})\mathbf{v} = \mathbf{u}^T\mathbf{v} = \mathbf{u}.\mathbf{v}$
- Any orthogonal set of vectors is linearly independent and hence is a basis spans the space.
- However there are some other properties of orthogonal set of vectors which makes them particularly desirable.

Orthogonality

What's in a basis?

- Let's use the following notation: For the Linear Transformation $\mathcal{R}^3 \rightarrow \mathcal{R}^2$

$$L\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{bmatrix} 3 & 4 & 5 \\ 2 & 2 & 1 \end{bmatrix} =$$

Orthogonality

What's in a basis?

- Let's use the following notation: For the Linear Transformation $\mathcal{R}^3 \rightarrow \mathcal{R}^2$

$$L\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{bmatrix} 3 & 4 & 5 \\ 2 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 3x + 4y + 5z \\ 2x + 2y + 1z \end{bmatrix}$$

- What's the basis for L?

Orthogonality

What's in a basis?

- Let's use the following notation: For the Linear Transformation $\mathcal{R}^3 \rightarrow \mathcal{R}^2$

$$L\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{bmatrix} 3\,4\,5 \\ 2\,2\,1 \end{bmatrix} = \begin{bmatrix} 3x + 4y + 5z \\ 2x + 2y + 1z \end{bmatrix}$$

- What's the basis for L? The basis is the number of linearly independent vectors in the matrix which are?

Orthogonality

What's in a basis?

- Let's use the following notation: For the Linear Transformation $\mathcal{R}^3 \rightarrow \mathcal{R}^2$
$$L\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{bmatrix} 3 & 4 & 5 \\ 2 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 3x + 4y + 5z \\ 2x + 2y + 1z \end{bmatrix}$$
- What's the basis for L? The basis is the number of linearly independent vectors in the matrix which are? $(3, 4, 5)^T$ and $(2, 2, 1)^T$
- What if we could do a similar Transformation with a different set of basis?

Orthogonality

What's in a basis?

- Let's use the following notation: For the Linear Transformation $\mathcal{R}^3 \rightarrow \mathcal{R}^2$
$$L\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{bmatrix} 3\,4\,5 \\ 2\,2\,1 \end{bmatrix} = \begin{bmatrix} 3x + 4y + 5z \\ 2x + 2y + 1z \end{bmatrix}$$
- What's the basis for L? The basis is the number of linearly independent vectors in the matrix which are? $(3, 4, 5)^T$ and $(2, 2, 1)^T$
- What if we could do a similar Transformation with a different set of basis?
$$L\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{bmatrix} 3 - 2\,4 - 2\,5 - 1 \\ 2 \quad 2 \quad 1 \end{bmatrix} = \begin{bmatrix} 3x - 2x + 4y - 2y + 5z - z \\ 2x + 2y + 1z \end{bmatrix} =$$
$$\begin{bmatrix} 3x - 2x + 4y - 2y + 5z - z \\ 2x + 2y + 1z \end{bmatrix} = \begin{bmatrix} x + 2y + 4z \\ 2x + 2y + 1z \end{bmatrix}$$

Orthogonality

What's in a basis?

- Let's use the following notation: For the Linear Transformation $\mathcal{R}^3 \rightarrow \mathcal{R}^2$
$$L\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{bmatrix} 3\,4\,5 \\ 2\,2\,1 \end{bmatrix} = \begin{bmatrix} 3x + 4y + 5z \\ 2x + 2y + 1z \end{bmatrix}$$
- What's the basis for L? The basis is the number of linearly independent vectors in the matrix which are? $(3, 4, 5)^T$ and $(2, 2, 1)^T$
- What if we could do a similar Transformation with a different set of basis?
$$L\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{bmatrix} 3 - 2 & 4 - 2 & 5 - 1 \\ 2 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 3x - 2x + 4y - 2y + 5z - z \\ 2x + 2y + 1z \end{bmatrix} =$$
$$\begin{bmatrix} 3x - 2x + 4y - 2y + 5z - z \\ 2x + 2y + 1z \end{bmatrix} = \begin{bmatrix} x + 2y + 4z \\ 2x + 2y + 1z \end{bmatrix}$$
- You may have seen these earlier as the task of reducing the matrices to *canonical form* or *echelon form*. Notice that for these matrices the computations would have been much easier! What does this remind you of?

Orthogonality

What's in a basis?

- Let's use the following notation: For the Linear Transformation $\mathcal{R}^3 \to \mathcal{R}^2$
$$L\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{bmatrix} 3\,4\,5 \\ 2\,2\,1 \end{bmatrix} = \begin{bmatrix} 3x + 4y + 5z \\ 2x + 2y + 1z \end{bmatrix}$$
- What's the basis for L? The basis is the number of linearly independent vectors in the matrix which are? $(3, 4, 5)^T$ and $(2, 2, 1)^T$
- What if we could do a similar Transformation with a different set of basis?
$$L\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{bmatrix} 3-2 & 4-2 & 5-1 \\ 2 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 3x - 2x + 4y - 2y + 5z - z \\ 2x + 2y + 1z \end{bmatrix} =$$
$$\begin{bmatrix} 3x - 2x + 4y - 2y + 5z - z \\ 2x + 2y + 1z \end{bmatrix} = \begin{bmatrix} x + 2y + 4z \\ 2x + 2y + 1z \end{bmatrix}$$
- You may have seen these earlier as the task of reducing the matrices to *canonical form* or *echelon form*. Notice that for these matrices the computations would have been much easier! What does this remind you of? The Gauss Jordan elimination of course!

Orthogonality

What's in a basis?

- Let's use the following notation: For the Linear Transformation $\mathcal{R}^3 \rightarrow \mathcal{R}^2$
$$L\begin{pmatrix}x\\y\\z\end{pmatrix} = \begin{bmatrix}3\,4\,5\\2\,2\,1\end{bmatrix} = \begin{bmatrix}3x+4y+5z\\2x+2y+1z\end{bmatrix}$$
- What's the basis for L? The basis is the number of linearly independent vectors in the matrix which are? $(3,4,5)^T$ and $(2,2,1)^T$
- What if we could do a similar Transformation with a different set of basis?
$$L\begin{pmatrix}x\\y\\z\end{pmatrix} = \begin{bmatrix}3-2\,4-2\,5-1\\2\quad\;\;2\quad\;\;1\end{bmatrix} = \begin{bmatrix}3x-2x+4y-2y+5z-z\\2x+2y+1z\end{bmatrix} =$$
$$\begin{bmatrix}3x-2x+4y-2y+5z-z\\2x+2y+1z\end{bmatrix} = \begin{bmatrix}x+2y+4z\\2x+2y+1z\end{bmatrix}$$
- You may have seen these earlier as the task of reducing the matrices to *canonical form* or *echelon form*. Notice that for these matrices the computations would have been much easier! What does this remind you of? The Gauss Jordan elimination of course!
- By doing such transformations we can reduce the Transformation to a *similar* Transformation.

Why Orthogonality?

Why do we need to understand all this?

Why Orthogonality?

Why do we need to understand all this?

- Because this is how almost all (high-dimensional) statistics is done. (More on that later.)
- If a vector is written in terms of orthogonal basis, the computations are easier.
  Say $\{\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_n}\}$ is a set of vectors constituting an orthogonal basis for ?

Why Orthogonality?

Why do we need to understand all this?

- Because this is how almost all (high-dimensional) statistics is done. (More on that later.)
- If a vector is written in terms of orthogonal basis, the computations are easier.
  Say $\{\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_n}\}$ is a set of vectors constituting an orthogonal basis for ?
- For any vector $\mathbf{x}' \in \mathcal{R}^n, \mathbf{x}' = a'_1\mathbf{x_1} + a'_2\mathbf{x_2} + ... + a'_n\mathbf{x_n}$
- And for any other vector $\mathbf{x}'' \in \mathcal{R}^n, \mathbf{x}'' = a''_1\mathbf{x_1} + a''_2\mathbf{x_2} + ... + a''_n\mathbf{x_n}$
- Now $\mathbf{x}'.\mathbf{x}'' = \mathbf{x}'^{\mathbf{T}}\mathbf{x}'' = (\sum a'_k\mathbf{x_1})^T(\sum a''_k\mathbf{x_1})$
- Now, what if the vectors $\mathbf{x_k}$ were not orthogonal?

Why Orthogonality?

Why do we need to understand all this?

- Because this is how almost all (high-dimensional) statistics is done. (More on that later.)
- If a vector is written in terms of orthogonal basis, the computations are easier.
  Say $\{\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_n}\}$ is a set of vectors constituting an orthogonal basis for ?
- For any vector $\mathbf{x}' \in \mathcal{R}^n, \mathbf{x}' = a'_1\mathbf{x_1} + a'_2\mathbf{x_2} + ... + a'_n\mathbf{x_n}$
- And for any other vector $\mathbf{x}'' \in \mathcal{R}^n, \mathbf{x}'' = a''_1\mathbf{x_1} + a''_2\mathbf{x_2} + ... + a''_n\mathbf{x_n}$
- Now $\mathbf{x}'.\mathbf{x}'' = \mathbf{x}'^{\mathbf{T}}\mathbf{x}'' = (\sum a'_k\mathbf{x_1})^T(\sum a''_k\mathbf{x_1})$
- Now, what if the vectors $\mathbf{x_k}$ were not orthogonal? Too many computations!
- Which is why orthogonality is a really handy concept.
- Another concept is *Orthonormal*. Where an Orthonormal basis is such that all vectors $\in \mathcal{B}$ are orthogonal and each $||\mathbf{x}|| = 1$.
- A matrix can also be orthonormal. What would that mean?

Diagonalization

Eigenvalues and Eigenvectors

- Eigenvalues are numbers which correspond to the solution of the equation $\mathbf{Av} = \lambda\mathbf{v}$ where $\mathbf{A}$ is a square matrix.
- Eigenvectors are the vectors which correspond to different values of λ
- We won't go into the details of how to determine these Eigenvalues and Eigenvectors however, all the Eigenvectors are orthogonal and thus, form a basis for the space.
- Eigenvalues and Eigenvectors give important information about the matrices.
- For a Linear Transformation $\mathbf{A} : \mathcal{R}^n \rightarrow \mathcal{R}^m, m \neq n$ the concept of Eigenvalues cannot exist. Instead we have the concept of *Singular Values*
- For an LT, $\mathbf{A} : \mathcal{R}^n \rightarrow \mathcal{R}^m, m \neq n$, Singular Value Decomposition is a factorization of the form $\mathbf{U}*$ where $\mathbf{U}$ is an $m \times m$ Unitary Matrix and $\mathbf{V}$ is an $n \times n$ Unitary Matrix.

References I

[1] Carl De Boor. *Applied Linear Algebra (Draft)*. ftp://ftp.cs.wisc.edu/Approx/book.ps. Accessed: 2017-12-02.