# COS 324, Precept #1:
# A Review of Linear Algebra

February 7, 2020

## 1 Introduction

This precept covers the basics of matrices and vectors, orthogonality, and subspaces. If you are already familiar with this material, you can skip this precept! We assume you have already encountered these topics in a linear algebra class: this precept is intended as a refresher, covering the main linear algebra ideas used in this class.

## 2 Definitions and Basics

A *vector* is an ordered set of numbers. For example, we can write

$$\mathbf{v} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

A *matrix* is an array of numbers, for example:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$

$A$ is a 2 by 3 matrix, since it has 2 rows and 3 columns. To specify parts of $A$, we use subscripts. $A_{i,j}$ is the number in the $i$-th row, $j$-th column of $A$. We call $A_{i,*}$ the $i$-th row of $A$, and $A_{*,j}$ the $j$-th column of $A$. For example:

$$A_{2,3} = 5, A_{1,*} = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}, A_{*,1} = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$$

Vectors are usually thought of as columns, but this can be inconvenient to write. Instead, we sometimes write $v^t = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}$ or $v = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}^t$, where $v^t$ is *$v$-transpose*. Transposing means to ”flip”: rows become columns, and columns become rows. For example,

$$A^t = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

The *dot product* of two vectors is found by summing the element-wise products of the vectors. We denote dot product as $v \cdot w$ or $v^t w$. For example, if $v = \begin{bmatrix}1 & 2 & 3\end{bmatrix}^t$ and $w = \begin{bmatrix}4 & 5 & 6\end{bmatrix}^t$.

$$v \cdot w = v^t w = 1 \cdot 4 + 2 \cdot 5 + 3 \cdot 6 = 32$$

To multiply matrices, find the dot product of the rows of the left matrix with the columns of the right matrix. For example, if

$$A = \begin{bmatrix}1 & 2\\ 3 & 4\end{bmatrix}, B = \begin{bmatrix}5 & 6\\ 7 & 8\end{bmatrix}$$

then we can write

$$AB = \begin{bmatrix}A_{1,*} \cdot B_{*,1} & A_{1,*} \cdot B_{*,2}\\ A_{2,*} \cdot B_{*,1} & A_{2,*} \cdot B_{*,2}\end{bmatrix} = \begin{bmatrix}19 & 22\\ 43 & 50\end{bmatrix}$$

# 3 Orthogonality

If $v \cdot w$ is positive, the vectors $v$ and $w$ are facing a similar direction. If $v \cdot w$ is negative, $v$ and $w$ are facing somewhat opposite directions. $v$ and $w$ are *orthogonal*, *normal*, or *perpendicular*, when their dot product is zero. For an example in two dimensions, see Figure 1.

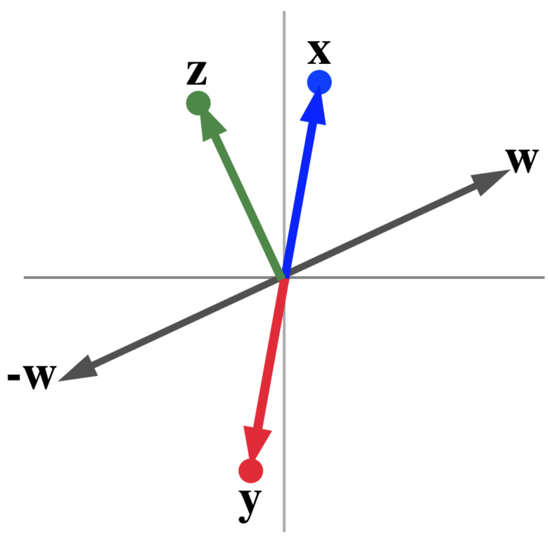


Figure 1: $z \cdot w = 0$, $x \cdot w > 0$, and $y \cdot w < 0$

# 4 Norm

The *norm* of a vector $v$, denoted $||v||$, is its length:

$$||v|| = \sqrt{\sum_{i=1}^{n} v_i^2}$$

where $v_i$ is the $i$-th element of $v$, and $n$ is the length of $v$.

# 5 Projection Onto a Vector

Given two vectors, $v$ and $w$, the *projection* of $v$ on $w$ is the vector closest to $v$, that is in the same direction as $w$. So, the projection of $v$ onto $w$ is always a multiple of $w$. The formula is:

$$proj_w(v) = \frac{v \cdot w}{||w||^2} w$$

Thus, the projection of $v$ on $w$ is closely related to $v \cdot w$. When $v \cdot w < 0$, $proj_w(v)$ lies in the direction of $-w$, and when $v \cdot w$, $proj_w(v)$ lies in the direction of $w$. For example, see Figure 2.

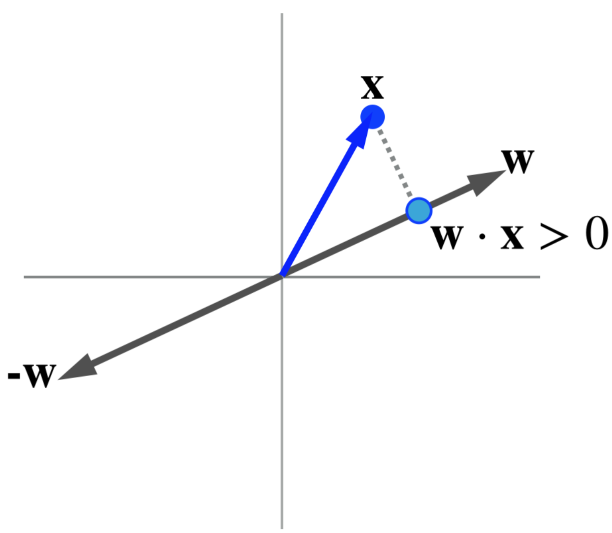


Figure 2: $w \cdot x > 0$, so $proj_w(x)$ (represented by the lighter dot) is in the direction of $w$.

Note that projection is simpler when $w$ is a unit vector (has length 1). When $||w|| = 1$:

$$proj_w(v) = (v \cdot w)w$$

Also, note that $proj_w(v) = 0$ if and only if $v$ is orthogonal to $w$. You can visualize this: when $v, w$ are perpendicular, $proj_w(v)$ vanishes.

# 6 Projection Onto a Subspace

A *linear combination* of a set of $k$ vectors $v_1, v_2, \ldots, v_k$ is some vector $a_1v_1 + a_2v_2 + \ldots + a_kv_k$ where $a_1, \ldots, a_k$ are scalar numbers. The *span* of some vectors $v_1, v_2, \ldots, v_k$ is a set of vectors, made up of all linear combinations of $v_1, \ldots, v_k$.

A *linear subspace* contains all linear combinations of some set of vectors: in other words, a linear subspace is a span of some set of vectors. The *dimension of a subspace* is the smallest number of vectors that would be needed to span that subspace. For example, any line through the origin is a 1-dimensional subspace, since a line consists of all linear combinations of a single vector. In three dimensional space ($\mathbb{R}^3$), any plane through the origin consists of all linear combinations of two vectors, so a plane through the origin is a 2-dimensional space.

In general, in $\mathbb{R}^d$, a subspace of $d-1$ dimensions is called a *hyperplane*.

In general, given any linear subspace you can find a set of orthogonal vectors whose linear combinations form the subspace. Assuming that you know a set of orthogonal vectors that span a subspace, you can easily project a vector $v$ onto the subspace. If your subspace $W$ is spanned by mutually orthogonal vectors $w_1, w_2, \ldots, w_k$, then the projection of $v$ can be found by summing the projection of $v$ onto each of the orthogonal vectors.

$$proj_W(v) = \sum_{i=1}^{k} proj_{w_i}(v) = \sum_{i=1}^{k} \frac{(v \cdot w_i)}{||w_i||^2} w_i$$

This is especially simple when $||w_i|| = 1$, since it simplifies to

$$proj_W(v) = \sum_{i=1}^{k} (v \cdot w_i) w_i$$

Just like projecting onto a vector, $proj_W(v) = 0$ exactly when $v$ is orthogonal to $W$ (that is, when $v$ is orthogonal to the vectors that span $W$). This leads to the next idea.

# 7 Perpendicular Spaces

Consider a linear subspace $W$. We call the set of vectors perpendicular to $W$, $W^\perp$ (pronounced $W$-perp). It turns out that $W^\perp$ is also a linear subspace. And the dimensions of $W$ and $W^\perp$ add up the dimension of the whole space!

For example, suppose $W$ is a 2-dimensional plane in $\mathbb{R}^3$. Then, $W^\perp$ will be a line perpendicular to $W$.

# 8 Application to Linear Classifiers

In $\mathbb{R}^d$, suppose $W$ is a 1-dimensional subspace (a line) formed by the vector $w$, where $w$ is the weight vector in our linear classifier. Then, as mentioned in lecture, $W^\perp$ will be a hyperplane, a subspace of $d-1$ dimensions.

Note that $W^\perp = \{x | w \cdot x = 0\}$ (the set of vectors $x$ that are orthogonal to $w$). $W^\perp$ is the boundary between the vectors $x$ where $w \cdot x > 0$ and the vectors $x$ where $w \cdot x < 0$.

Thus, thinking about $W^\perp$ gives us a nice spacial interpretation of a linear classifier. If $w$ is our weight vector, the $d-1$-dimensional hyperplane $W^\perp$ separates our positively-classified examples from our negatively-classified examples. Changing our weight vector $w$ simply changes the orientation of $W^\perp$.